

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2004, 19:48:00 ; Search time 4444.98 Seconds  
(without alignments)  
16274.436 Million cell updates/sec

Title: US-10-080-960-1

Perfect score: 1669  
Sequence: 1 caccgctcgcctcgcctcgcct.....aaaaaaaaaaaaaaaaaaaaa 1669

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hng:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_to:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hng:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_scs:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hng\_hum:\*  
31: em\_hng\_inv:\*  
32: em\_hng\_other:\*  
33: em\_hng\_mus:\*  
34: em\_hng\_pln:\*  
35: em\_hng\_rod:\*  
36: em\_hng\_mam:\*  
37: em\_hng\_vrt:\*  
38: em\_sy:\*  
39: em\_hngo\_hum:\*  
40: em\_hngo\_mus:\*  
41: em\_hngo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID        | Description        |
|------------|--------|-------------|--------|--------------|--------------------|
| 1          | 1669   | 100.0       | 1669   | 6 AX592444   | AX592444 Sequence  |
| 2          | 1506   | 90.2        | 3610   | 6 BC063462   | BC063462 Homo sapi |
| 3          | 1461   | 87.5        | 1461   | 6 AX592446   | AX592446 Sequence  |
| 4          | 1386.6 | 83.1        | 2312   | 6 HSA582015  | AJ582015 Homo sapi |
| 5          | 1383.6 | 82.8        | 2675   | 6 HSA431184  | AJ431184 Homo sapi |
| 6          | 1133   | 67.9        | 1673   | 9 HSA535838  | AJ535838 Homo sapi |
| 7          | 1132.4 | 67.8        | 1623   | 9 HSA512465  | AJ512465 Homo sapi |
| 8          | 1084.4 | 65.0        | 1651   | 9 HSA535839  | AJ535839 Homo sapi |
| 9          | 1011.8 | 60.7        | 1389   | 4 AY344579   | AY344579 Bos tauru |
| 10         | 838.8  | 50.3        | 104120 | 9 AC067838   | AC067838 Homo sapi |
| 11         | 838.8  | 50.3        | 185872 | 9 AC091144   | AC091144 Homo sapi |
| 12         | 799.6  | 47.9        | 2874   | 10 BC030474  | BC030474 Mus muscu |
| 13         | 799.6  | 47.9        | 2978   | 10 BC062213  | BC062213 Mus muscu |
| 14         | 661    | 39.6        | 1657   | 5 GGA535692  | AJ535692 Gallus ga |
| 15         | 626    | 37.5        | 191072 | 2 AC115775   | AC115775 Mus muscu |
| 16         | 581.2  | 34.8        | 597    | 6 AX314120   | AX314120 Sequence  |
| 17         | 507.6  | 30.4        | 1333   | 10 BC022579  | BC022579 Mus muscu |
| 18         | 374    | 22.4        | 185912 | 2 BX640592   | BX640592 Danio rer |
| 19         | 370.8  | 22.2        | 165781 | 2 BX321906   | BX321906 Homo sapi |
| 20         | 309.2  | 18.5        | 2027   | 9 AK056199   | AK056199 Homo sapi |
| 21         | 300    | 18.0        | 1114   | 9 BC004884   | BC004884 Homo sapi |
| 22         | 298    | 17.9        | 1550   | 6 AX817125   | AX817125 Sequence  |
| 23         | 288    | 17.3        | 2180   | 9 AK098505   | AK098505 Homo sapi |
| 24         | 257.4  | 15.4        | 450    | 6 AR420976   | AR420976 Sequence  |
| 25         | 257.4  | 15.4        | 450    | 6 BD116529   | BD116529 EST and e |
| 26         | 257.4  | 15.4        | 450    | 6 AC102943   | AC102943 Homo sapi |
| 27         | 257.4  | 15.4        | 66715  | 2 AC102943   | AC102943 Homo sapi |
| 28         | 257.4  | 15.4        | 66715  | 2 AC102943   | AC102943 Homo sapi |
| 29         | 212.4  | 14.1        | 1506   | 5 AJ606069   | AJ606069 Fugu rubr |
| 30         | 212.4  | 12.7        | 1866   | 5 CH153752   | AJ53752 Gallus ga  |
| 31         | 198.2  | 11.9        | 1542   | 10 RNC035753 | AJ53753 Rattus no  |
| 32         | 184    | 11.0        | 1714   | 3 AK114811   | AK114811 Clona int |
| 33         | 183.8  | 11.0        | 3211   | 10 MMT542393 | AJ542393 Mus muscu |
| 34         | 180.8  | 10.8        | 1475   | 6 AX356085   | AX356085 Sequence  |
| 35         | 180.8  | 10.8        | 2069   | 9 BC036037   | BC036037 Homo sapi |
| 36         | 180.8  | 10.8        | 2557   | 6 AX356083   | AX356083 Sequence  |
| 37         | 179    | 10.7        | 194879 | 2 AC109163   | AC109163 Mus muscu |
| 38         | 178.4  | 10.7        | 1335   | 3 DME302046  | AJ302046 Drosophi  |
| 39         | 176.8  | 10.6        | 1335   | 3 AY061932   | AY061932 Drosophi  |
| 40         | 166    | 9.9         | 2408   | 6 AX833606   | AX833606 Sequence  |
| 41         | 166    | 9.9         | 2408   | 6 AK095482   | AK095482 Homo sapi |
| 42         | 125.4  | 7.8         | 181589 | 5 AC146544   | AC146544 Gasterost |
| 43         | 125.4  | 7.8         | 197879 | 5 AC146543   | AC146543 Gasterost |
| 44         | 116.2  | 7.0         | 218747 | 10 AC121599  | AC121599 Mus muscu |
| 45         | 109.8  | 6.6         | 249360 | 2 AC127920   | AC127920 Rattus no |

#### ALIGNMENTS

RESULT 1  
AX592444  
LOCUS AX592444 1669 bp DNA linear PAT 27-JAN-2003  
DEFINITION Sequence 1 from Patent W00232962.  
ACCESSION AX592444  
VERSION AX592444.1 GI:27950546  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS  
TITLE  
Glucksmann, M.A. and Meyers, R.  
80090, 52874, 52880, 63497, and 33425 methods and compositions of  
human proteins and uses thereof

JOURNAL Patent: WO 023962-A 1 25-APR-2002;  
 Millennium Pharmaceuticals, Inc. (US)  
 FEATURES Location/Qualifiers  
 source 1..1669  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 163..1623  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAD62017.1"  
 /db\_xref="GI:27950547"  
 /db\_xref="REVIEW:CAD62017"  
 /translation="MKVTPPGQVTDSDMQCENDQWPLSTRSSHIKEVVELGKFER  
 KEFKSSLDQGHKMEAPTHLNSFLKKEGLTPNKRKWEIDSYIMLWMSPLTGEIG  
 RLGGCCADACFTINRNTYLHHMTKAPFYGTDFNIDSLPLPRKAHDMVFEESPK  
 NNYKLFHKRVLTLFNTATFSSHSLPLTOYLSIEVLKSLRYLPLDSKNLRLKSL  
 APLVYVQSDCPDSDSDSVRELMYIEVDSYGELEKRLDPOOLKNPASMADAGPYR  
 IIAQYKFIILAFENACDDYITEKFWRLKGLVYPIYYSBSITTHMUSNKSAILIYSEF  
 SHERELASTIRLSDSDRLYEAVYEWKUGETISNQLRTALREKRWGVQNDQNDYD  
 APECVVCTKVMANIRLOEKGLPPKRWELADHLSCPEPTVAFSPILRTPPLSLREWM  
 ISSFEQSKKEAQLRMLVDNRNQNFSSEFQWGLVFKD"

ORIGIN

Query Match 100.0%; Score 1669; DB 6; Length 1669;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACGCGTCGCTCTGCTCTAGTGTGACTTTGGCGCTCAGGTGATCCATGACTTTT 60  
 Db 1 CACGCGTCGCTCTGCTCTAGTGTGACTTTGGCGCTCAGGTGATCCATGACTTTT 60  
 QY 61 TAAAGCCAATATATTTCTTCTCTCTGAGTGTGCTGCTGCTTCTTCTTCTTCTT 120  
 Db 61 TAAAGCCAATATATTTCTTCTCTCTGAGTGTGCTGCTGCTTCTTCTTCTTCTT 120  
 QY 121 TTTTCTTCTTCTTCTTCTTCTTCTGATACCGTGTGAGTCTAATGAAAGTCAAGGCGCT 180  
 Db 121 TTTTCTTCTTCTTCTTCTTCTTCTGATACCGTGTGAGTCTAATGAAAGTCAAGGCGCT 180  
 QY 181 CCCCAGGAGTTACAGACTCCATGCATGCTTCAATGATCAGTGGCTTTTCTTCAAC 240  
 Db 181 CCCCAGGAGTTACAGACTCCATGCATGCTTCAATGATCAGTGGCTTTTCTTCAAC 240  
 QY 241 AGGAGCAGGAGACATTAAGAGGTCATGTTGCTGGGGAGTTGAAAGGAGAG 300  
 Db 241 AGGAGCAGGAGACATTAAGAGGTCATGTTGCTGGGGAGTTGAAAGGAGAG 300  
 QY 301 TTTTAAAGTTCAGTTTGCAGATGACATCAAAAATGAGGAGACCTTACGATCTT 360  
 Db 301 TTTTAAAGTTCAGTTTGCAGATGACATCAAAAATGAGGAGACCTTACGATCTT 360  
 QY 361 AATTCTTTCTTAAAGAAAGAGATTGACCTTCAACAGAGAAAGAAATGGAAATTGAC 420  
 Db 361 AATTCTTTCTTAAAGAAAGAGATTGACCTTCAACAGAGAAAGAAATGGAAATTGAC 420  
 QY 421 AGCTACCCCATTAATGCTGTGTGTCGCCGCTGACGAGGAGACCTGGAGGTTTGGCAA 480  
 Db 421 AGCTACCCCATTAATGCTGTGTGTCGCCGCTGACGAGGAGACCTGGAGGTTTGGCAA 480  
 QY 481 TGTGAGCAGATGCTGTGTCTTCAACATCAACCGGACCTTCTTCATCATCATGACC 540  
 Db 481 TGTGAGCAGATGCTGTGTCTTCAACATCAACCGGACCTTCTTCATCATCATGACC 540  
 QY 541 AAAGCATCTCTCTCTATGCTGCTGATCTTAAAGATGATGATGATGATGATGATGAT 600  
 Db 541 AAAGCATCTCTCTCTATGCTGCTGATCTTAAAGATGATGATGATGATGATGATGAT 600  
 QY 601 GCCCATCATGATGCTGGCTGTTTTCATGAAGAGTCCCGGAAACCAATTAAGCTCTTT 660  
 Db 601 GCCCATCATGATGCTGGCTGTTTTCATGAAGAGTCCCGGAAACCAATTAAGCTCTTT 660  
 QY 661 CATAAACGAGTATTAACCTGTTCAACTACACTGCCAGTTCACAGAGCAATTCACACTTG 720

Db 661 CATAAACGAGTATTAACCTGTTCAACTACACTGCCAGTTCACAGAGCAATTCACACTTG 720  
 QY 721 CCACATCACTACCCAACTACTTGAAGAGCATTTGAAGTCTGAAAGTCACTCCATCTT 780  
 Db 721 CCACATCACTACCCAACTACTTGAAGAGCATTTGAAGTCTGAAAGTCACTCCATCTT 780  
 QY 781 CCTTTGCACTCCAAACCAAGCTTGAAGAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
 Db 781 CCTTTGCACTCCAAACCAAGCTTGAAGAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
 QY 841 GACTGTGACCCCACTCAGACAGGACAGCTATGTTGCGAGCTGATGATCTTACATCGAG 900  
 Db 841 GACTGTGACCCCACTCAGACAGGACAGCTATGTTGCGAGCTGATGATCTTACATCGAG 900  
 QY 901 GTGCAATCTATGCTGAATGTTTACGAAACAAAGACCTCCCTCAGCAGCTGAAATCC 960  
 Db 901 GTGCAATCTATGCTGAATGTTTACGAAACAAAGACCTCCCTCAGCAGCTGAAATCC 960  
 QY 961 GCTCTTATGATGCTCGATGCTTTTATGATCATTCACAGATTAAGTTATCTTACT 1020  
 Db 961 GCTCTTATGATGCTCGATGCTTTTATGATCATTCACAGATTAAGTTATCTTACT 1020  
 QY 1021 TTTGAGATGAGATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 Db 1021 TTTGAGATGAGATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 QY 1081 GGGGTAGTCCCTGTATATTAAGATCCCGCAGCATCAAGACTGCTTCCAGTAA 1140  
 Db 1081 GGGGTAGTCCCTGTATATTAAGATCCCGCAGCATCAAGACTGCTTCCAGTAA 1140  
 QY 1141 AGTGTATTTCTGTATGATGAAATTTTCTCACCCCGAGGAACTGGCAATGATCATCA 1200  
 Db 1141 AGTGTATTTCTGTATGATGAAATTTTCTCACCCCGAGGAACTGGCAATGATCATCA 1200  
 QY 1201 CTGATTTCTGATGACAGATTTGTATGAGGCTATGATGATGATGATGATGATGATGAT 1260  
 Db 1201 CTGATTTCTGATGACAGATTTGTATGAGGCTATGATGATGATGATGATGATGATGAT 1260  
 QY 1261 TCTAACCGGACATTTCTGACAGCTCTCAGGGAACGAAATGGGAGTGCAGACGCTAAC 1320  
 Db 1261 TCTAACCGGACATTTCTGACAGCTCTCAGGGAACGAAATGGGAGTGCAGACGCTAAC 1320  
 QY 1321 CAGACATTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
 Db 1321 CAGACATTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
 QY 1381 AGCTTCAAGAAAGGGCTTACACCCCAAGATGGAGGACAGAAATCCACCTGAGT 1440  
 Db 1381 AGCTTCAAGAAAGGGCTTACACCCCAAGATGGAGGACAGAAATCCACCTGAGT 1440  
 QY 1441 TGCCAGAGGCCACAGGTTTGTCTTCTCACCTCCGAGCTCACCTTTGAGCTCTTGG 1500  
 Db 1441 TGCCAGAGGCCACAGGTTTGTCTTCTCACCTCCGAGCTCACCTTTGAGCTCTTGG 1500  
 QY 1501 CGAGAGATGATGATTTTCAGCTTTGACCAATCCAGAAAGAACCCAGGACATTAAGTGG 1560  
 Db 1501 CGAGAGATGATGATTTTCAGCTTTGACCAATCCAGAAAGAACCCAGGACATTAAGTGG 1560  
 QY 1561 CTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
 Db 1561 CTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
 QY 1621 TGATTTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1669  
 Db 1621 TGATTTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1669

RESULT 2  
 BC063462 3610 bp mRNA linear PRI 11-DEC-2003  
 LOCUS  
 DEFINITION Homo sapiens cDNA clone MGC:74714 IMAGE:5198060, complete cds.  
 ACCESSION BC063462

VERSION B063462.1 GI:39645816  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 3610)  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Bhat, N.K., Alech, S.F., Zeeberg, B., Buetow, K.H., Mang, J., Hsieh, F., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Dlatenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stopleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Wierley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyak, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 2 (bases 1 to 3610)  
 Strausberg, R.  
 Direct Submission  
 Submitted (08-DEC-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: http://www.nisc.nih.gov/  
 Contact: nisc.mgc@nih.gov  
 Akheri, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granito, S., Guan, X., Gupta, D., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Masello, C., Masker, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stancipop, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wechter, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://imgc.lnl.gov  
 Series: IRK Plate: 135 Row: f Column: 2  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 1992647.  
 Location/Qualifiers  
 1. 3610  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:74714 IMAGE:5198060"  
 /tissue\_type="Brain, adult, 6 pooled whole brains"  
 /clone\_lib="NIH MGC\_114"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 271. 1836  
 /codon\_start=1

ORIGIN  
 Query Match 90.2%; Score 1506; DB 9; Length 3610;  
 Best Local Similarity 98.7%; Pred. No. 0;  
 Matches 1518; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
 /product="Unknown (protein for MGC:74714)"  
 /protein\_id="AAH63462.1"  
 /db\_xref="GI:39645817"  
 /translation="MVRIRKRLKSLCLCTATVFLVTLQALDTVENIMKTGPQG  
 VTSNQCENPDRPLSNTRSEHKEVMVLEGFERKEFSSSLQDGHTEAPLHNM  
 SPLKKGLEFNRKWELEDSYPLMWSPLTETGCGCADAFETINRTEAFSRH  
 TKPLGTPENRPLDPLRKAHDMVAHESPKNMYLFPKPYITLNTYATFSR  
 SHPLPTQIESLEVKSLRKYPLRLOSKKRLKPLAVYQSDDPEDSDYREL  
 MTIYDVSGECKRNDLPOQLNPNASNDPFRYIAKXKTLIPENAVDDYITEK  
 FMPKLGAVPYVYGSITDMLPNKSKALVSESRPRELSSYTRRLSDRLIEAY  
 VEKMLKEISNRLTLAVEPKWQVQNDYIDAFEKCVTKWANRLDEKGLP  
 KRMEADETHLSCEPFTVFAFSPILRPLPLSLREMTISFEOSKEKQALRLVDNRN  
 FSSGEFMGLVFXD"  
 643. 1467  
 /note="Glyco transf. 10; Region: Glycosyltransferase family  
 10 (Fucosyltransferase). This family of  
 fucosyltransferases are the enzymes transferring fucose  
 from GDP-fucose to GlcNAc in an alpha1,3 linkage. This  
 family is known as glycosyltransferase family 10"  
 /db\_xref="CD:pfam0852"

118 TTTTCTTTTCTTTCTTTTGGCCCTGGATACCGTGGAGATCTAATGAAGTCACGGGC 177  
 |||||  
 331 TTTCCTGCTGCACTCCAGGCTTGATACCCTTGAGATCTAATGAAGTCACGGGC 390  
 |||||  
 178 CCTCCCGAGGAGTTACAGACTCCATGCTTCAATGATCAATGAGCTTATCTTAC 237  
 |||||  
 391 COTCCCGAGGAGTTACAGACTCCATGCTTCAATGATCAATGAGCTTATCTTAC 450  
 |||||  
 238 ACCAGGACACCGAGACATTAAGAGTCAATGAGTGGGGAGTTGAAGAG 297  
 |||||  
 451 ACCAGGACACCGAGACATTAAGAGTCAATGAGTGGGGAGTTGAAGAG 510  
 |||||  
 298 GAGTTAAAGTTCAGTTGCAAGATGACATACAAAATGAGAGACCTACGCAT 357  
 |||||  
 511 GAGTTAAAGTTCAGTTGCAAGATGACATACAAAATGAGAGACCTACGCAT 570  
 |||||  
 358 CTTAATTAATTTCTTAAGAAAGAGATTGACCTTCAACAGAGAAATGGAATTG 417  
 |||||  
 571 CTTAATTAATTTCTTAAGAAAGAGATTGACCTTCAACAGAGAAATGGAATTG 630  
 |||||  
 418 GACAGCTACCCCATANGCTGTGGTGTCGCCGCTGACGGGGAGACTGGAGGTTAGGC 477  
 |||||  
 631 GACAGCTACCCCATANGCTGTGGTGTCGCCGCTGACGGGGAGACTGGAGGTTAGGC 690  
 |||||  
 478 CATGTGAGAGAGATGCTGTTTCTTCAACATCAACCGAGCTTACCTCATCATCATG 537  
 |||||  
 691 CATGTGAGAGAGATGCTGTTTCTTCAACATCAACCGAGCTTACCTCATCATCATG 750  
 |||||  
 538 ACCAAGATTCCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 597  
 |||||  
 751 ACCAAGATTCCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 810  
 |||||  
 598 AAAGCCATCATGACTGAGGCTGTTTTCATGAGAGTCCCGGAAATCAATTAATAGCTC 657  
 |||||  
 811 AAAGCCATCATGACTGAGGCTGTTTTCATGAGAGTCCCGGAAATCAATTAATAGCTC 870  
 |||||  
 658 TTTCATTAACAGATATTAATCTTGTTCACATCACTGACAGGTTCAAGAGGATTCAC 717  
 |||||  
 871 TTTCATTAACAGATATTAATCTTGTTCACATCACTGACAGGTTCAAGAGGATTCAC 930  
 |||||  
 718 TTGGCACTACATACCATATCTTGAGAGCATTTGAAGTCTGAAGTCACTCCGATACCTA 777  
 |||||  
 931 TTGGCACTACATACCATATCTTGAGAGCATTTGAAGTCTGAAGTCACTCCGATACCTA 990  
 |||||  
 778 GTTCCTTTGACATCCAAATCAAGCTTGAAGAGATTTGCTCCGCTGGTGTATGATACG 837  
 |||||  
 991 GTTCCTTTGACATCCAAATCAAGCTTGAAGAGATTTGCTCCGCTGGTGTATGATACG 1050  
 |||||

|            |  |      |   |                            |
|------------|--|------|---|----------------------------|
| OY         |  | 898  | TCAGACCTGTGACCACCATATCAACAGGACAGCTATGTTCCGGAGCTGAATGACTTACATC     | 897                        |
| Db         |  | 1051 | TGAGCTGTGACCCACCATCAAGAGGACAGCTATGTTCCGGAGCTGAATGACTTACATC        | 111                        |
| OY         |  | 898  | GAGGTGCATTCTTAGTGATGTTTACGAAACAAGAAGCTCCCTCAGCAGCTAATAAT          | 957                        |
| Db         |  | 1111 | GAGGTGCATTCTTAGTGATGTTTACGAAACAAGAAGCTCCCTCAGCAGCTAATAAT          | 117                        |
| OY         |  | 958  | CCAGCTCTATGATGCGCATGGCTTTTATAGATCATTTGCCAGTATTAAGTTATCTTA         | 1011                       |
| Db         |  | 1171 | CCAGCTCTATGATGCGCATGGCTTTTATAGATCATTTGCCAGTATTAAGTTATCTTA         | 123                        |
| OY         |  | 1018 | GCTTTTGAGATGACAGTTTGTATGACATCATCTGAGAAAGTTCTGAGAGCCACTGAAA        | 107                        |
| Db         |  | 1231 | GCTTTTGAGATGACAGTTTGTATGACATCATCTGAGAAAGTTCTGAGAGCCACTGAAA        | 129                        |
| OY         |  | 1078 | CTGGSGGTAGTCCCTGTATATTACGATCCCCACATCACAGACTGGCTTCCAAGTAA          | 113                        |
| Db         |  | 1291 | CTGGSGGTAGTCCCTGTATATTACGATCCCCACATCACAGACTGGCTTCCAAGTAA          | 135                        |
| OY         |  | 1138 | AAAAAGTCTATTCTTGATCAGAAATTTCTACCCCGAGAGAACCTGGAGATTACATCAG        | 119                        |
| Db         |  | 1391 | AAAAAGTCTATTCTTGATCAGAAATTTCTACCCCGAGAGAACCTGGAGATTATTCAG         | 141                        |
| OY         |  | 1198 | CGACTGATTTCTGATGACAGATTGTATGAGGCTTATGTAAATGAAAGCTGAAGGCTGAG       | 125                        |
| Db         |  | 1411 | CGACTGATTTCTGATGACAGATTGTATGAGGCTTATGTAAATGAAAGCTGAAGGCTGAG       | 147                        |
| OY         |  | 1258 | ATCTCTAACAGAGGACCTTCGACAGCTCAGAGGAAACGGAATGGGAGTGCAGAACCTTC       | 131                        |
| Db         |  | 1471 | ATCTCTAACAGAGGACCTTCGACAGCTCAGAGGAAACGGAATGGGAGTGCAGAACCTTC       | 153                        |
| OY         |  | 1318 | AACCAAGACAAATTACATCGATGCATTGTAGTGTATGATGTGTGCACCAAGTGTGGCTAAT     | 137                        |
| Db         |  | 1531 | AACCAAGACAAATTACATCGATGCATTGTAGTGTATGATGTGTGCACCAAGTGTGGCTAAT     | 159                        |
| OY         |  | 1378 | ATCAGGCTTAGGAAAGAGGCTTACACCCCAAAGATGGAGGACAGAAAGTATCCCACCTG       | 143                        |
| Db         |  | 1591 | ATCAGGCTTAGGAAAGAGGCTTACACCCCAAAGATGGAGGACAGAAAGTATCCCACCTG       | 165                        |
| OY         |  | 1438 | AGTTGCCAGAGCCCAACAGTGTTCCTTTCTCACCACTCCGGACTCCACTCTTGAAGCTCT      | 149                        |
| Db         |  | 1651 | AGTTGCCAGAGCCCAACAGTGTTCCTTTCTCACCACTCCGGACTCCACTCTTGAAGCTCT      | 171                        |
| OY         |  | 1498 | TTGGAGAGAGATGAGGATTTCCAGCTTTGAACAATCCAAAGAAAGCCAGGACACTAAG        | 155                        |
| Db         |  | 1711 | TTGGAGAGAGATGAGGATTTCCAGCTTTGAACAATCCAAAGAAAGCCAGGACACTAAG        | 177                        |
| OY         |  | 1558 | TGGCTGTGATGATGAAATCAAACTTTTCATCTCAGAGTTTTGGGCGCTAGATTTCAAG        | 161                        |
| Db         |  | 1771 | TGGCTGTGATGATGAAATCAAACTTTTCATCTCAGAGTTTTGGGCGCTAGATTTCAAG        | 183                        |
| OY         |  | 1618 | GACTGATTTCAAAAATGATCGAAATGGAACAGAAAAA                             | 165                        |
| Db         |  | 1831 | GACTGATTTCAAAAATGATCGAAATGGAACAGAAAAA                             | 186                        |
| RESULT 3   |  |      |   |                            |
| AX592446   |  |      | 1461 bp   | DNA linear PAT 27-JAN-2003 |
| LOCUS      |  |      |   |                            |
| DEFINITION |  |      | Sequence 3 from Patent WO0232962.                                 |                            |
| ACCESSION  |  |      | AX592446  |                            |
| VERSION    |  |      | AX592446.1  | GI:27950548                |
| KEYWORDS   |  |      |   |                            |
| SOURCE     |  |      | Homo sapiens (human)  |                            |
| ORGANISM   |  |      | Homo sapiens  |                            |
| REFERENCE  |  |      | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |                            |
| AUTHORS    |  |      | Mammalia; Duteria; Primates; Catarrhini; Homindae; Homo.          |                            |
| TITLE      |  |      | Glucksmann, M.A. and Meyers, R.                                   |                            |
|            |  |      | 80090, 52874, 52880, 63497, and 33425 methods and compositions of |                            |
|            |  |      | human proteins and uses thereof                                   |                            |

|                            |   |
|----------------------------|---|
| JOURNAL                    | Patent: WO 0232962-A 3 25-APR-2002,   |
| FEATURES                   | Millennium Pharmaceuticals, Inc. (US)   |
| source                     | Location/Qualifiers<br>1..1461<br>/organism="Homo sapiens"<br>/mol_type="unassigned DNA"<br>/db_xref="taxon:9606" |
| ORIGIN                     |   |
| Query Match                | 87.5%; Score 1461; DB 6; Length 1461;   |
| Best Local Similarity      | 100.0%; Pred. No. 0;  |
| Matches 1461; Conservative | 0; Mismatches 0; Indels 0; Gaps 0;  |
| QY                         | 163 ATGAAGGTACGGGCGCCCTCCCAGGAGATTAAAGACTCCATGCATGCTTCATATATCAG 222   |
| DB                         | 1 ATGAAGGTACGGGCGCCCTCCCAGGAGATTAAAGACTCCATGCATGCATCATATCAG 60  |
| QY                         | 223 TGGCCTTTTCTTAACACGACGAGACGCACTAAAAAGGTGATGGTTAGCTGGGG 282   |
| DB                         | 61 TGCCCTTTTCTTAACACGACGAGACGCACTAAAAAGGTGATGGTTAGCTGGGG 120  |
| QY                         | 283 AGCTTGAAGAAGAGAGTTTAAAAGTTCAGTTTGCAAGATGACATACAAAAATGAG 342   |
| DB                         | 121 AAGTTGAAGAAGAGAGTTTAAAAGTTCAGTTTGCAAGATGACATACAAAAATGAG 180   |
| QY                         | 343 GAAGACCCAGCATCTTAATTCATATTTCTTAAGAAAGAAAGATTGACCTTCAACAGGAA 402   |
| DB                         | 181 GAAGACCTTAGCCATCTTAATTCATATTTCTTAAGAAAGAAAGATTGACCTTCAACAGGAA 240   |
| QY                         | 403 AGAAATGGAATTGAGACGCTACCCTTAATGCTCTGAGTCCCCGCTGACGGGGAG 462  |
| DB                         | 241 AGAAATGGAATTGAGACGCTACCCTTAATGCTCTGAGTCCCCGCTGACGGGGAG 300  |
| QY                         | 463 ACTGGGAGGTATGGCCAATGTGAGACAATGCTGTTTTCTTCACCATCACCGAECTAC 522   |
| DB                         | 301 ACTGGGAGGTATGGCCAATGTGAGACAATGCTGTTTTCTTCACCATCACCGAECTAC 360   |
| QY                         | 523 CTCCTCATCATACATACCAAAGCAATCTCTCTATGATGATGACTTAACATATGATGC 582   |
| DB                         | 361 CTCCTCATCATATACCAAAGCAATCTCTCTATGATGATGACTTAACATATGATGC 420   |
| QY                         | 583 TTACTCTGCCCTGGAAAGCCCATCATGACTGGGCTGTTTTTCATGAAGATCCCCGAA 642   |
| DB                         | 421 TTACTCTGCCCTGGAAAGCCCATCATGACTGGGCTGTTTTTCATGAAGATCCCCGAA 480   |
| QY                         | 643 AACATTTATAAGCTCTTTTATATAACCAAGTATACCTTGTCATCACTACCTGCACAGTTC 702  |
| DB                         | 481 AACATTTATAAGCTCTTTTATATAACCAAGTATACCTTGTCATCACTACCTGCACAGTTC 540  |
| QY                         | 703 AGCAGGCAATCCCACTTECCCACTAACCCANACTTGAGAGCATTTGAAGTCTGAAG 762  |
| DB                         | 541 AGCAGGCAATCCCACTTECCCACTAACCCANACTTGAGAGCATTTGAAGTCTGAAG 600  |
| QY                         | 763 TCACCTCCGATACCTAGTTCCTTTGCACTCCAAAAACAAGCTTAGAAAAAGACTTGCCTCC 822   |
| DB                         | 601 TCACCTCCGATACCTAGTTCCTTTGCACTCCAAAAACAAGCTTAGAAAAAGACTTGCCTCC 660   |
| QY                         | 823 CTGGTGATATGACAGTCAAGCTGACCCCACTCAGACAGGGGACAGCTATGTTCCGGAG 882  |
| DB                         | 661 CTGGTGATATGACAGTCAAGCTGACCCCACTCAGACAGGGGACAGCTATGTTCCGGAG 720  |
| QY                         | 883 CTGATGACTTACATCGAGGTGATTCCTATGTGTAATGTTTACGAAACAAAGACTTCCT 942  |
| DB                         | 721 CTGATGACTTACATCGAGGTGATTCCTATGTGTAATGTTTACGAAACAAAGACTTCCT 780  |
| QY                         | 943 CAGAGGCTGAAGAAATCCAGGCTCATATGATGAGCGGCTTTATATAGATCATTTGCACAG 1002   |
| DB                         | 781 CAGAGGCTGAAGAAATCCAGGCTCATATGATGAGCGGCTTTATATAGATCATTTGCACAG 840  |
| QY                         | 1003 TATAAGTTTATCCTAGCTTTTGAAGATGAGCTTTGTGATGACTACATCACTCGAAGTTC 1062   |
| DB                         | 841 TATAAGTTTATCCTAGCTTTTGAAGATGAGCTTTGTGATGACTACATCACTCGAAGTTC 900   |



QY 1063 TGGAGGCACTGAAACCTGGGGGTAGTCCCTGTATATTACGATCCGCCAGATCAACAGC 1122  
Db TGGAGGCACTGAAACCTGGGGGTAGTCCCTGTATATTACGATCCGCCAGATCAACAGC 960  
QY 1123 TGGCTTCCAACTAACAAAGTGTATTCTGTATCAAGATTTTCTCACCCCGAGGAACG 1182  
Db TGGCTTCCAACTAACAAAGTGTATTCTGTATCAAGATTTTCTCACCCCGAGGAACG 1020  
QY 1183 GCMACTTCATCAACACAGCTGGATTCTGATGACAGATTGATGAGGCTTATGTAAATGG 1242  
Db GCMACTTCATCAACACAGCTGGATTCTGATGACAGATTGATGAGGCTTATGTAAATGG 1080  
QY 1243 AAGCTGAAGGGTGAAGATCTTAACCCAGCACTTCTGACAGCTCTCAAGGAACGGAATGG 1302  
Db AAGCTGAAGGGTGAAGATCTTAACCCAGCACTTCTGACAGCTCTCAAGGAACGGAATGG 1140  
QY 1303 GGAATGCAAGAGCTCAACAGCAATTAATCATGATCATTTGATGATGTGTGCACC 1362  
Db GGAATGCAAGAGCTCAACAGCAATTAATCATGATCATTTGATGATGTGTGCACC 1200  
QY 1363 AAGGTGTGGGCTAATATATAGGCTTACAGAAAGGGCTTACACCCAAAGATGGAGGGA 1422  
Db AAGGTGTGGGCTAATATATAGGCTTACAGAAAGGGCTTACACCCAAAGATGGAGGGA 1260  
QY 1423 GAAGATACCACTGATGTTGCCAGAGCCCACTGTTTCTTCCACCACTCCGAGCT 1482  
Db GAAGATACCACTGATGTTGCCAGAGCCCACTGTTTCTTCCACCACTCCGAGCT 1320  
QY 1483 CCACCTTTGAGCTCTTTCGAGAGATGTGATTTCCAGCTTTGAACATCCAGAAAGAA 1542  
Db CCACCTTTGAGCTCTTTCGAGAGATGTGATTTCCAGCTTTGAACATCCAGAAAGAA 1380  
QY 1543 GCCCAGGACATAGGCTGCTGTGATAGATCAAACTTTTCACTCAAGAGTTTGG 1602  
Db GCCCAGGACATAGGCTGCTGTGATAGATCAAACTTTTCACTCAAGAGTTTGG 1440  
QY 1603 GGCTTAGTATTCAGAGCTGA 1623  
Db GGCTTAGTATTCAGAGCTGA 1461  
RESULT 4  
LOCUS HSA582015 2312 bp mRNA linear PRI 04-SEP-2003  
DEFINITION Homo sapiens mRNA for alpha3-fucosyltransferase (FUT10 gene).  
ACCESSION AJ582015.1 GI:34481985  
VERSION alpha3-fucosyltransferase; FUT10 gene.  
KEYWORDS Homo sapiens (human).  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 Martinez-Duncker I., Candellier J.J., Oriol R. and Mollicone R.  
AUTHORS Cloning, expression and genomic organization of two new human  
TITLE alpha3-fucosyltransferases (FUT10 and FUT11)  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2312)  
AUTHORS Oriol R.  
TITLE Direct Submission  
JOURNAL Submitted (03-SEP-2003) Oriol R., U504, Inserm, 16 Av. Paul  
Vaillant-Couturier, 94807, FRANCE  
COMMENT Variants: AJ512465, AJ535839, AJ535838, AJ431184.  
FEATURES  
source  
1..2312  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/map="BP11.23"  
/clone="479"  
1..2312  
/gene="FUT10"

## CDS

357..1796  
/gene="FUT10"  
/function="adds fucose in alpha3-linkage to GlcNAc"  
/codon\_start=1  
/evidence="experimental"  
/product="alpha3-fucosyltransferase"  
/protein\_id="CAE46499.1"  
/db\_xref="GI:34481986"  
/translation="MVRIRPKILASCLCVTAIVPLVLTQVMEIGKEERKEPSSS  
LQGHITMBEAPTHLNSFLKESLTPNRKKELEDSYPIMLMWSPLTGGTGLGCGA  
DAPFTINRYLHNMTKAFLEFGDIFNIDSLPEPKAHMDVAHRESFKNNYLFH  
KPVITLFNRYATPSRSHLELTQYLESLEVLKSLRVLPLQSKKRLRLPLVYQ  
SDCDPSDRSVYREIMTYIEVDYSGECLRNKDLPOQLNPAISMADGYRILIAQV  
ILAFENACDYLTEKFRPLKLGVPYVYGSITDMLPSNKSALIVSEFSELA  
SVYRLSDDRDLVEAVEMKLEISNORLALREKXGVODVNDVYIDAFECVVC  
TKYMANTRLOEKGLPKRWEADPHLSCEPEVPFAPSPTRTPPLSLRWMLSPEQS  
KKEQALRLVLVDNQNFSOEFNGLVFKD"  
polyA\_signal  
2274..2279  
/gene="FUT10"  
polyA\_site  
2295  
/gene="FUT10"

## ORIGIN

Query Match 83.1%; Score 1386.6; DB 9; Length 2312;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1389; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 263 AGGTGATGGTTGAGCTGGGAGCTTGAAGAGAGGTTTAAAGTTCCAGTTTSCAG 322  
Db AGGTGATGGTTGAGCTGGGAGCTTGAAGAGAGGTTTAAAGTTCCAGTTTSCAG 495  
QY 323 ATGACATACAAAATGAGAGAAACCTACGATCTTAATTCATTCTTAAAGAAAG 382  
Db ATGACATACAAAATGAGAGAAACCTACGATCTTAATTCATTCTTAAAGAAAG 555  
QY 383 GATTGACCTTCAACAGAAAGAAATGGAAATGGAACGTAACCCATTAATGCTGTGG 442  
Db GATTGACCTTCAACAGAAAGAAATGGAAATGGAACGTAACCCATTAATGCTGTGG 615  
QY 443 GGTCCCGCTGACGGGGAGAGTGGAGGTTTAAAGCAATGTGAGAGATGCTTGTTC 502  
Db GGTCCCGCTGACGGGGAGAGTGGAGGTTTAAAGCAATGTGAGAGATGCTTGTTC 675  
QY 503 TCACATCAACCGGAGCTTACCTCATATACATGACCAAGATTCCTTTATAGGTA 562  
Db TCACATCAACCGGAGCTTACCTCATATACATGACCAAGATTCCTTTATAGGTA 735  
QY 563 CTGACTTTAATAGATAGTACTTGCCTCGAAGGCCATCATGACTGGGCTGTTT 622  
Db CTGACTTTAATAGATAGTACTTGCCTCGAAGGCCATCATGACTGGGCTGTTT 795  
QY 623 TTCAATGAAGTCCCGGAAAACAATTAAGCTCTTTCATTAACAAGATTAATCTTGT 682  
Db TTCAATGAAGTCCCGGAAAACAATTAAGCTCTTTCATTAACAAGATTAATCTTGT 855  
QY 796 TTCAATGAAGTCCCGGAAAACAATTAAGCTCTTTCATTAACAAGATTAATCTTGT 855  
Db TTCAATGAAGTCCCGGAAAACAATTAAGCTCTTTCATTAACAAGATTAATCTTGT 922  
QY 856 TCAACTACATGCGACAGCTTACAGAGGATCCGACTTCCACATCACTACCAATACCTGG 915  
Db TCAACTACATGCGACAGCTTACAGAGGATCCGACTTCCACATCACTACCAATACCTGG 982  
QY 915 TTAAGAAAAGACTTGCCTCGGTGTATATACATGATGAGCTTGAAGCCCAATGAGAC 962  
Db TTAAGAAAAGACTTGCCTCGGTGTATATACATGATGAGCTTGAAGCCCAATGAGAC 1035  
QY 962 TTAAGAAAAGACTTGCCTCGGTGTATATACATGATGAGCTTGAAGCCCAATGAGAC 962  
Db TTAAGAAAAGACTTGCCTCGGTGTATATACATGATGAGCTTGAAGCCCAATGAGAC 1035  
QY 1035 GAGACAGCTATGTTGCGAGCTGATGATACATGAGAGTGGATTCCTTAAGTGAATGTT 922  
Db GAGACAGCTATGTTGCGAGCTGATGATACATGAGAGTGGATTCCTTAAGTGAATGTT 1095  
QY 1095 GAGACAGCTATGTTGCGAGCTGATGATACATGAGAGTGGATTCCTTAAGTGAATGTT 922  
Db GAGACAGCTATGTTGCGAGCTGATGATACATGAGAGTGGATTCCTTAAGTGAATGTT 1095  
QY 923 TAGGAACAAAGACTTCCCTGAGAGCTGAAAATCCAGCTTATGAGATCCGATGCTT 962  
Db TAGGAACAAAGACTTCCCTGAGAGCTGAAAATCCAGCTTATGAGATCCGATGCTT 962

|            |      |  |   |
|------------|------|--|---|
| Db         | 1096 | TACGAAACAAAGATCCCTTCAGACAGGTGAAATAATCCAGCTCTATGAGATGCCAGATGGCT | 1155  |
| Qy         | 983  | TTTATAGATCATTTGCACAGTAATTAAGTTTATCTTACCTTTTGGAGATGCATGTTTGATG  | 1042  |
| Db         | 1156 | TTTATAGATCATTTGCACAGTAATTAAGTTTATCTTACCTTTTGGAGATGCATGTTTGATG  | 1215  |
| Qy         | 1043 | ACTACATCACTGAGAAAGTTTGGAGGCCACTGAAACTGGGGAGTACCTCTATATTATAG    | 1102  |
| Db         | 1216 | ACTACATCACTGAGAAAGTTTGGAGGCCACTGAAACTGGGGAGTACCTCTATATTATAG    | 1275  |
| Qy         | 1103 | GATCCCCCAGCATCAAGACTGGCTTCCAGATTAACAAAGTGTCTATTCTTGTATCCAGAT   | 1162  |
| Db         | 1276 | GATCCCCCAGCATCAAGACTGGCTTCCAGATTAACAAAGTGTCTATTCTTGTATCCAGAT   | 1335  |
| Qy         | 1163 | TTTTCACACCCAGGAACTGGCAGTATACATCAAGACATGAGATTCCTGATGACATTTGT    | 1222  |
| Db         | 1336 | TTTTCACACCCAGGAACTGGCAGTATACATCAAGACATGAGATTCCTGATGACATTTGT    | 1395  |
| Qy         | 1223 | ATGAGGCTTATGTAAGATGGAAGCTGAAGGCTGAGATCTCTTAACAGCAGCTTCTGACAG   | 1282  |
| Db         | 1396 | ATGAGGCTTATGTAAGATGGAAGCTGAAGGCTGAGATCTCTTAACAGCAGCTTCTGACAG   | 1455  |
| Qy         | 1283 | CTTCACAGGAAACGGAATTTGGGAGTGCATAACCTGCAACCGACATTTACATGATGAT     | 1342  |
| Db         | 1456 | CTTCACAGGAAACGGAATTTGGGAGTGCATAACCTGCAACCGACATTTACATGATGAT     | 1515  |
| Qy         | 1343 | TTGAGTGTATGTGTGACACCAAGGTGTGGCTTAATATCAGGCTTCAGAAAAGGCTTAC     | 1402  |
| Db         | 1516 | TTGAGTGTATGTGTGACACCAAGGTGTGGCTTAATATCAGGCTTCAGAAAAGGCTTAC     | 1575  |
| Qy         | 1403 | CACCCAAAAGATGGGAGCGAGAAATATCCCACTGAGTTGCCAGACCCCAAGTGTG        | 1462  |
| Db         | 1576 | CACCCAAAAGATGGGAGCGAGAAATATCCCACTGAGTTGCCAGACCCCAAGTGTG        | 1635  |
| Qy         | 1463 | CTTCTCACCTCCGAGCTCCACCTTGTAGCTCTTTGGAGAGATGTGGATTTCCAGCT       | 1522  |
| Db         | 1636 | CTTCTCACCTCCGAGCTCCACCTTGTAGCTCTTTGGAGAGATGTGGATTTCCAGCT       | 1695  |
| Qy         | 1523 | TTGACCAATCCAGAAAAGAGCCGAGCACTTAAGTGTGCTTGTATGAAATCAAAACT       | 1582  |
| Db         | 1696 | TTGACCAATCCAGAAAAGAGCCGAGCACTTAAGTGTGCTTGTATGAAATCAAAACT       | 1755  |
| Qy         | 1583 | TTTATATCTCAAGATTTTGGGGCTAGTATTCAGAGATGATTTTCAAAATGATCAGAT      | 1642  |
| Db         | 1756 | TTTATATCTCAAGATTTTGGGGCTAGTATTCAGAGATGATTTTCAAAATGATCAGAT      | 1815  |
| Qy         | 1643 | GAAGACAGAAAAA 1655   |   |
| Db         | 1816 | GAAGACAGCTAGA 1828   |   |
| RESULT 5   |      |  |   |
| HSAA431184 |      | 2675 bp  | mRNA linear PRI 26-FEB-2002   |
| LOCUS      |      |  |   |
| DEFINITION |      |  | Homo sapiens partial mRNA for putative alpha 1,3-fucosyl transferase (FUT10 gene).  |
| ACCESSION  |      |  | AF431184.1 GI:18996309  |
| VERSION    |      |  | FUT10 gene; putative alpha 1,3-fucosyl transferase.   |
| KEYWORDS   |      |  | Homo sapiens (human)  |
| SOURCE     |      |  | Homo sapiens  |
| ORGANISM   |      |  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.   |
| REFERENCE  |      |  | 1 Roos,C., Kolmer,M., Mattila,P. and Renkonen,R. Composition of Drosophila melanogaster proteome involved in fucosylated glycan metabolism J. Biol. Chem. 277 (5), 3168-3175 (2002) |
| JOURNAL    |      |  |   |
| MEDLINE    |      |  | 21671316  |
| PUBMED     |      |  | 11696403  |
| REFERENCE  |      |  | 2 (bases 1 to 2675)   |
| AUTHORS    |      |  | Roos,C.   |
| TITLE      |      |  | Direct Submission   |

Db 724 TTGGAGAGCATTTGAAGTCTGTAAGTCACTCCGATACCTAGTTCCTTTGGAGTCCAAAAAC 783  
QY 799 AAGCTTAGAAAAAGACTGTCCTCCGTGCTGTAATGATACAGTACGTGTGAGCCACCATCA 858  
Db 784 AAGCTTAGAAAAAGCTTCTCCGTGCTGTAATGATACAGTACGTGTGAGCCACCATCA 843  
QY 859 GACAGGAGACGCTATGTTCCGAGAGTGAATGATTAATGAGTGTGATTCCTTATGATGAA 918  
Db 844 GACAGGAGACGCTATGTTCCGAGAGTGAATGATTAATGAGTGTGATTCCTTATGATGAA 903  
QY 919 TGTGTTACGAAAACAAAGACTCCCTCCAGAGTGAATAATCCAGCTCTATGATGCCGAT 978  
Db 904 TGTGTTACGAAAACAAAGACTCCCTCCAGAGTGAATAATCCAGCTCTATGATGCCGAT 963  
QY 979 GGCCTTTATAGGATCATTGACAGATTAAGTTATTCCTAGCTTTTGAAGATGAGATTGT 1038  
Db 964 GGCCTTTATAGGATCATTGACAGATTAAGTTATTCCTAGCTTTTGAAGATGAGATTGT 1023  
QY 1039 GATGACTACATACATGAGAAAGTTCTGAGGCGCACTGAAACTGGGGGTAATCCTCTATAT 1098  
Db 1024 TATGACTACATACATGAGAAAGTTCTGAGGCGCACTGAAACTGGGGGTAATCCTCTATAT 1083  
QY 1099 TACGAGTCCCCAGAGATCAGACAGTGGCTTCCAAATAAATAAGTCTATTCCTGATATCA 1158  
Db 1084 TACGAGTCCCCAGAGATCAGACAGTGGCTTCCAAATAAATAAGTCTATTCCTGATATCA 1143  
QY 1159 GAATTTTCTACCCAGGAAACTGGCAAGTTATCATCAAGACATGATGATGATGACAG 1218  
Db 1144 GAATTTTCTACCCAGGAAACTGGCAAGTTATCATCAAGACATGATGATGATGACAG 1203  
QY 1229 TTGTAATGAGGCTATGTAAGATGAGAGTGAAGGCTGAGATCTCTAACCGAGACTTCG 1278  
Db 1204 TTGTAATGAGGCTATGTAAGATGAGAGTGAAGGCTGAGATCTCTAACCGAGACTTCG 1263  
QY 1279 ACAGCTCTCAGGAAACGAAATGGGAGTGCACAGCTCAACAGCAATTAATCATGAT 1338  
Db 1264 ACAGCTCTCAGGAAACGAAATGGGAGTGCACAGCTCAACAGCAATTAATCATGAT 1323  
QY 1339 GCATTTGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398  
Db 1324 GCATTTGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1383  
QY 1399 TTACCAACCCAAAGATGGGAGGAGAGATATCCACCTGAGTTCCCGAGGCCACAGTG 1458  
Db 1384 TTACCAACCCAAAGATGGGAGGAGAGATATCCACCTGAGTTCCCGAGGCCACAGTG 1443  
QY 1459 TTTGCTTTCTCACTCCGACTTCACCTTTGAGCTTTTGCAGAGATGAGATTTC 1518  
Db 1444 TTTGCTTTCTCACTCCGACTTCACCTTTGAGCTTTTGCAGAGATGAGATTTC 1503  
QY 1519 ACCTTTGAACAATCCAAAGAAAGAGCCAGGACATAGAGTGGCTGATAGGAATCA 1578  
Db 1504 ACCTTTGAACAATCCAAAGAAAGAGCCAGGACATAGAGTGGCTGATAGGAATCA 1563  
QY 1579 AACTTTTATCTCAAGAGTTTGGGCTAGTATTTCAAGACTGATTTCAAAATGATCA 1638  
Db 1564 AACTTTTATCTCAAGAGTTTGGGCTAGTATTTCAAGACTGATTTCAAAATGATCA 1623  
QY 1639 GAATGAACAGAAAAA 1655  
Db 1624 GAATGAACAGACTGA 1640

RESULT 6  
HSA535838 1673 bp mRNA linear PRI 01-FEB-2003  
LOCUS HSA535838 1673 bp mRNA linear PRI 01-FEB-2003  
DEFINITION Homo sapiens mRNA for putative alpha3-fucosyltransferase (FUT10  
gene), splice variant 1.  
ACCESSION AJ535838  
VERSION AJ535838.1 GI:27475656  
KEYWORDS alpha3-fucosyltransferase; alternative splicing; FUT10 gene.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Candelier, J.J., Martinez-Duncker, I., Oriol, R. and Molliecore, R.  
TITLE Cloning expression and genomic organization of two new human  
alpha3-fucosyltransferases (FUT10 and FUT11)  
JOURNAL unpublished  
REFERENCE 2 (bases 1 to 1673)  
AUTHORS Oriol, R.  
TITLE Direct Submission  
JOURNAL Submitted (27-DEC-2002) Oriol R., U504, INSERM, 16 Avenue Paul  
Valliant-Couratier, 94807, Villejuif, FRANCE  
COMMENT related sequences: AJ512465, AJ535839.  
FEATURES  
source  
1. 1673  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1. 1673  
/gene="FUT10"  
/function="adds fucose via alpha 1-3 linkage to GlcNAc"  
/note="splice variant 1, lacks the transmembrane domain"  
/codon\_start=1  
/product="putative alpha3-fucosyltransferase"  
/protein\_id="GI:27475657"  
/db\_xref="GI:27475657"  
/db\_xref="GO:081417"  
/db\_xref="SPTRXBL:081417"  
/translation="MTYDGGKPERKERSSSLDQHTKMEAPHLNLSPLKKEGLPFR  
KRKHELDSPIMLMWSFLTSETRLOCCGADAFPTLNRTLYLHHHTKAPLPICGDN  
IDSLPLERKHHPDQAVHEBSPRONTKLPHKPVITLFTNATFSRSHSLPLTQYIES  
IEVILSLRYVPLDQSNKRLKRLAPLVYVSDDDPSDSDSVRELMYILEVDSYEC  
LRNEDLQOLKPNASMDADGFRILIAQYFILLFENAVDDDYTEKFRPLKGVVAV  
YVGSPTIDMLPSKSHLIVSESHPELAASVIRLSDSDRLYEAVVEMKLEISNO  
RLTALBERKMGVQDVNVQYIDAFEMVCTKTMANIRIQEKSKSKGVIEPAGWPS"

ORIGIN  
Query Match 67.9%; Score 1133; DB 9; Length 1673;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1136; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 256 AATAAAGAGTCATGCTTGGAGCTGGGGAGTTGAAAGAGAGTTTAAAGTCCAGT 315  
Db 467 AATATGAAAGTCATGCTTGGAGCTGGGGAGTTGAAAGAGAGTTTAAAGTCCAGT 526  
QY 316 TTGCAGATGACATACAAAATGAGAGAGCACTACGATCTTAATTCATTTCTTAAG 375  
Db 527 TTGCAGATGACATACAAAATGAGAGAGCACTACGATCTTAATTCATTTCTTAAG 586  
QY 376 AAAGAAGATTGACCTTCAACAGGAAAGAAATGGGAATTGACAGCTTACCCATTATG 435  
Db 587 AAAGAAGATTGACCTTCAACAGGAAAGAAATGGGAATTGACAGCTTACCCATTATG 646  
QY 436 CTCTGAGTGTCCCGCTGACGAGGGGAGACTGGAGGTTAGGCCAATGAGAGAGATGCT 495  
Db 647 CTCTGAGTGTCCCGCTGACGAGGGGAGACTGGAGGTTAGGCCAATGAGAGATGCT 706  
QY 496 TGTGTTTTCACCATCAACCGGACCTTACCTTCATCATCACTGACCAAGATCTCTCTTC 555  
Db 707 TGTGTTTTCACCATCAACCGGACCTTACCTTCATCATCACTGACCAAGATCTCTCTTC 766  
QY 556 TATGTAATCTGACTTTAATCATAGATAGCTTACCTGCTGCGAAGAGCCATCATGACTGG 615  
Db 767 TATGTAATCTGACTTTAATCATAGATAGCTTACCTGCTGCGAAGAGCCATCATGACTGG 826  
QY 616 GCTGTTTTCATGAGAGTCCCGGAAAAACAAATTAAGCTCTTTCATTAACCAAGATGAT 675  
Db 827 GCTGTTTTCATGAGAGTCCCGGAAAAACAAATTAAGCTCTTTCATTAACCAAGATGAT 886  
QY 676 AACTTGTTCACATCACTGCCACCTTGACAGAGCATTTCCCACTTGGCCACTAATACCA 735

|    |      |  |      |
|----|------|--|------|
| Db | 887  | ACCTGTTCAACTACACTCCACGTTCCAGCAGGACATCCCACTGGCCACTACTACCA   | 946  |
| Qy | 736  | TACTTGAGAGCATTAAGTCTGAAGTCACTCCCATACATAGTTCCTTGAGTCCAA     | 795  |
| Db | 947  | TACTTGAGAGCATTAAGTCTGAAGTCACTCCCATACATAGTTCCTTGAGTCCAA     | 1006 |
| Qy | 796  | AACAAGCTTAGAAAAAGACTGCTCCGCTGCTGTATGTAACATCACTGTGACCA      | 855  |
| Db | 1007 | AACAAGCTTAGAAAAAGACTGCTCCGCTGCTGTATGTAACATCACTGTGACCA      | 1066 |
| Qy | 856  | TCAGACAGGAGCAGCATGTTGGCAGTGAATCTCATCGAGGTGATTCCTATAGT      | 915  |
| Db | 1067 | TCAGACAGGAGCAGCATGTTGGCAGTGAATCTCATCGAGGTGATTCCTATAGT      | 1126 |
| Qy | 916  | GAATGTTAGAAACAAAGACCTCCCTCAGCAGCTGAAAAATCCAGCCTCATGATGC    | 975  |
| Db | 1127 | GAATGTTAGAAACAAAGACCTCCCTCAGCAGCTGAAAAATCCAGCCTCATGATGC    | 1186 |
| Qy | 976  | GATGCTTTTATGAGTCACTGAGATTAAGTTCCTATAGTTCCTATAGTTCCTATAGT   | 1035 |
| Db | 1187 | GATGCTTTTATGAGTCACTGAGATTAAGTTCCTATAGTTCCTATAGTTCCTATAGT   | 1246 |
| Qy | 1036 | TGTGATGACTACATCACTGAGAAATCTGGAGCCACTGAAACTGGGGGTAGTCCCTGTA | 1095 |
| Db | 1247 | TGTGATGACTACATCACTGAGAAATCTGGAGCCACTGAAACTGGGGGTAGTCCCTGTA | 1306 |
| Qy | 1096 | TATTACGGATCCCCCGCATCAAGATCTGGCTTCCAAAGTAAACAAAGTCTATCTGTA  | 1155 |
| Db | 1307 | TATTACGGATCCCCCGCATCAAGATCTGGCTTCCAAAGTAAACAAAGTCTATCTGTA  | 1366 |
| Qy | 1156 | TCAGAAATTTCTACCCCGAGGAACTGGCAATCTCATCGACGCTGATTCGATGAC     | 1215 |
| Db | 1367 | TCAGAAATTTCTACCCCGAGGAACTGGCAATCTCATCGACGCTGATTCGATGAC     | 1426 |
| Qy | 1216 | AGATTGTATGAGGCTATGTAGATGAAAGTGAAGGTTGAGATCTTAACACGACTT     | 1275 |
| Db | 1427 | AGATTGTATGAGGCTATGTAGATGAAAGTGAAGGTTGAGATCTTAACACGACTT     | 1486 |
| Qy | 1276 | CTACACAGCTCTCGGGAAGGAAATGGGAGTGCAGACGTCACACAGCAATTAATATC   | 1335 |
| Db | 1487 | CTACACAGCTCTCGGGAAGGAAATGGGAGTGCAGACGTCACACAGCAATTAATATC   | 1546 |
| Qy | 1336 | GATGCAATTTGAGTGTATGTGTGACCAAGTGTGGGCTAAATATCAGGCTTCAAGAAAG | 1395 |
| Db | 1547 | GATGCAATTTGAGTGTATGTGTGACCAAGTGTGGGCTAAATATCAGGCTTCAAGAAAG | 1606 |
| Qy | 1396 | G 1396   |      |
| Db | 1607 | G 1607   |      |

RESULT 7  
 HSA512465 1623 bp mRNA linear PRI 02-JAN-2003  
 LOCUS Homo sapiens mRNA for putative alpha 1,3-fucosyl transferase (FUT10 gene).  
 ACCESSION AU512465  
 VERSION AU512465.1 GI:24370972  
 KEYWORDS alpha 1,3-fucosyl transferase; FUT10 gene.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1  
 AUTHORS Candelier, J., Martinez-Duncker, I., Oriol, R., and Mollison, R.  
 TITLE Cloning and genomic organization of a new alpha3-fucosyltransferase  
 JOURNAL FUT10  
 REFERENCE 2  
 AUTHORS Oriol, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-OCT-2002) Oriol R., US04, INSEPM, 16 Av. Paul

|                            |   |
|----------------------------|---|
| COMMENT                    | Valiant-Coururier, 94807 Villejuif, FRANCE  |
| FEATURES                   | related sequences: AU515838, AU515839.  |
| source                     | 1..1623<br>Location/Qualifiers<br>/organism="Homo sapiens"<br>/mol_type="mRNA"<br>/db_xref="taxon:9606"<br>/chromosome="8"<br>/map="Spl1.23"<br>/dev_stage="40 to 70 day embryo"<br>/country="France"<br>1..1623<br>/gene="FUT10"<br>345..1604<br>/gene="FUT10"<br>/codon_start=1<br>/evidence=experimental<br>/product="putative alpha 1,3-fucosyl transferase"<br>/protein_id="CAD54669.1"<br>/db_xref="GI:24370973"<br>/db_xref="GOA:O81VJ3"<br>/db_xref="SPTREMBL:O81VJ3"<br>/translation="MVRIOQRKLASCLVATVPLLVTLQVWVLEKTEREPKSS<br>LQDHTKMEAPHLNLSFLKEGLTENRKRKWLDSYPTLMWSPLTGTGLGQGA<br>DACEFTINRIVLHHTMKAPLFYGDPIIDSLPRAKHDVAVFEESPKNNYKLFH<br>KPVITLFRVATFERSHSLPLTOYLESIVLSKLVYVLCRNLRLRLAPLVYQ<br>SDCPSPDRDSYVLEMTYIEVDYSGCELRNKLPOOLKXNPASMDADGFRYIAQYK<br>ILAEFNACDXYITREKFRPLKIGVYVYVYSGSITDMLPSNKSATLVSEPSHPELA<br>SYIRLSDDELVAIVEMKLGKISINQRLITLRRKRKGVQVDVQNDNYIDAPECHVC<br>TKWMANRLQEKVSKSVGLEPAGWPS" |
| gene                       |   |
| CDS                        |   |
| ORIGIN                     |   |
| Query Match                | 67.8%; Score 1132.4; DB 9; Length 1623;   |
| Best Local Similarity      | 99.9%; Pred. No. 0;   |
| Matches 1133; Conservative | 0; Mismatches 1; Indels 0; Gaps 0;  |
| Qy                         | 263 AGGTGATGTTGAGCTGGGGAAGTTTGAAGAAGAGTTTAAAGTTCAGTTGCAAG 322   |
| Db                         | 424 AGGTGATGTTGAGCTGGGGAAGTTTGAAGAAGAGTTTAAAGTTCAGTTGCAAG 483   |
| Qy                         | 323 ATGACATACAAAATGAGAGAGACCTACCTATTAATTCATTTCTTAAGAAAG 382   |
| Db                         | 484 ATGACATACAAAATGAGAGAGACCTACCTATTAATTCATTTCTTAAGAAAG 543   |
| Qy                         | 383 GATTGACCTTCAAGAGAAAGAAATGGAAATGACAGCTACCCCATTTAGCTGTGT 442  |
| Db                         | 544 GATTGACCTTCAAGAGAAAGAAATGGAAATGACAGCTACCCCATTTAGCTGTGT 603  |
| Qy                         | 443 GGTCCCGCTGACGGGGGAGACTGGAGGTTTAAAGCCAAATGTGAGAGATGCTTTTCT 502   |
| Db                         | 604 GGTCCCGCTGACGGGGGAGACTGGAGGTTTAAAGCCAAATGTGAGAGATGCTTTTCT 663   |
| Qy                         | 503 TCACCATACACCGGACCTACCTCATCATGACCAAGATTCCTTCTATAGTA 562  |
| Db                         | 664 TCACCATACACCGGACCTACCTCATCATGACCAAGATTCCTTCTATAGTA 723  |
| Qy                         | 563 CTGACTTTAATATAGTATGCTTACCTTGCTGGGAAAGCCATCATGACTGGGCTGTT 622  |
| Db                         | 724 CTGACTTTAATATAGTATGCTTACCTTGCTGGGAAAGCCATCATGACTGGGCTGTT 783  |
| Qy                         | 623 TTCATGAGAGTCCCGGAAACAAATTAAGCTCTTTCATTAACAGATGATTAACCTTGT 682   |
| Db                         | 784 TTCATGAGAGTCCCGGAAACAAATTAAGCTCTTTCATTAACAGATGATTAACCTTGT 843   |
| Qy                         | 683 TCACCTACATGCGACGTTGACGAGGATTCCTCCATTCGACCTAACCTTACTTGG 742  |
| Db                         | 844 TCACCTACATGCGACGTTGACGAGGATTCCTCCATTCGACCTAACCTTACTTGG 903  |
| Qy                         | 743 AGAGCATGAAGTCTGAAGCACTCGATACCTAGTTCCTTTCAGAGTCCAAAG 802   |
| Db                         | 904 AGAGCATGAAGTCTGAAGCACTCGATACCTAGTTCCTTTCAGAGTCCAAAG 963   |
| Qy                         | 803 TTGAAAAAGACTGCTCCGCTGCTGTATGATGACGACGACTGACCCACATCAGACA 862   |

Db 964 TTGAAAAAGAGCTGCTGCGTGTGTATGTAAGTCAGCTGAGCTGAGCCACACACACACACA 1023  
QY 863 GGGACAGCTATGTTGCGGAGCTGATGACTTACATGAGAGTGATGCCATGAGTGAATGTT 922  
Db 1024 GGGACAGCTATGTTGCGGAGCTGATGACTTACATGAGAGTGATGCCATGAGTGAATGTT 1083  
QY 923 TACGAAACAAAGACCTCCCTCAGAGCTGAAAAATCCAGCTCTATGATGCCAGTGGCT 982  
Db 1084 TACGAAACAAAGACCTCCCTCAGAGCTGAAAAATCCAGCTCTATGATGCCAGTGGCT 1143  
QY 983 TTTATGAGTCAATGCAAGTATAGTTTATCCATGCTTTTGAAGATGAGAGTGGATG 1042  
Db 1144 TTTATGAGTCAATGCAAGTATAGTTTATCCATGCTTTTGAAGATGAGAGTGGATG 1203  
QY 1043 ACTACATCACTGAGAAGTTCTGAGGCCACATGAAATCGGGGGTAGTCCCTGATATTAACG 1102  
Db 1204 ACTACATCACTGAGAAGTTCTGAGGCCACATGAAATCGGGGGTAGTCCCTGATATTAACG 1263  
QY 1103 GATCCCCCAGCTACACAGTGTGGCTTCCAGTAACAAAGTGTATTTCTGTATCAGAAAT 1162  
Db 1264 GATCCCCCAGCTACACAGTGTGGCTTCCAGTAACAAAGTGTATTTCTGTATCAGAAAT 1323  
QY 1163 TTTCTCACCCCGAGGAAGCTGCAAGTTACATCAGACGACTGGATTCTGATGACGATTTG 1222  
Db 1324 TTTCTCACCCCGAGGAAGCTGCAAGTTACATCAGACGACTGGATTCTGATGACGATTTG 1383  
QY 1223 ATGAGGCTATGTAGATGAAAGCTGAAAGGTGAGATCTTAACACAGCACTTCTGACAG 1282  
Db 1384 ATGAGGCTATGTAGATGAAAGCTGAAAGGTGAGATCTTAACACAGCACTTCTGACAG 1443  
QY 1283 CTCTCAGGGAACGGAATGGGGAGTGCAGACGTCACACGAGCAATTAATGATGATGCAT 1342  
Db 1444 CTCTCAGGGAACGGAATGGGGAGTGCAGACGTCACACGAGCAATTAATGATGATGCAT 1503  
QY 1343 TTGAGTGTATGTGTGTGACCAAGGTGTGGCTAATATCAGGCTTCAGAAAAAG 1396  
Db 1504 TTGAGTGTATGTGTGTGACCAAGGTGTGGCTAATATCAGGCTTCAGAAAAAG 1557

RESULT 8  
LOCUS HSAS35839 1651 bp mRNA linear PRI 01-FEB-2003  
DEFINITION Homo sapiens mRNA for putative alpha3-fucosyltransferase (FUT10 gene), inactive short splice variant.  
ACCESSION AJ535839  
VERSION AJ535839.1 GI:27475658  
KEYWORDS alpha3-fucosyltransferase; alternative splicing; FUT10 gene.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Candelier,J.J., Martinez-Duncker,I., Oriol,R. and Mollicone,R.  
TITLE Cloning expression and genomic organization of two new human alpha3-fucosyltransferases (FUT10 and FUT11)  
JOURNAL Unpublished  
AUTHORS Oriol,R.  
TITLE 2 (bases 1 to 1651)  
REFERENCE Direct Submission  
JOURNAL Submitted (27-DEC-2002) Oriol R., US04, INSERM, 16 Avenue Paul  
TITLE Vallant-Couturier, 94807, Villejuif, FRANCE  
COMMENT related sequences: AJ512465, AJ535838.  
FEATURES  
source 1..1651  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1..1651  
/gene="FUT10"  
559..1632  
/gene="FUT10"  
/function="unknown"

/note="inactive short splice variant, lacks the transmembrane domain and the first and half of the second conserved peptide motifs"  
/codon\_start=1  
/product="putative alpha3-fucosyltransferase"  
/protein\_id="CAD59772.1"  
/db\_xref="GI:27475659"  
/db\_xref="GOA:O81V16"  
/db\_xref="SPRMBL:O81V16"  
/translation="MGIGLPYALVYPADGDMVYRPMWRCLPLIHOPDLPPSSHD OSILWMSOESPLAENDGTDPRNIDSLP.PRAHDMVAFHESPNMVKLPHKVI TLNVTATFSSHSLPLITTOYLESIEVLSKRLPLVPLQSNKRLKRLAIVYQSDCD PPSDRSVRELMLTVLEVDYGECLRNKDPQQLKDPASVDADGFRIILAQYFIIAF ENAVCDYIIEKFWPKLGVPVYGPSITLMDPSNKSALIVSEFSSHRELA5YIR RLSDSDRLYEAYVEMKLGKISINQRLTLALREKMGVDVQNDNYIDAEFCWCTRYW ANIRLEKVSXSKVGIKIPAMPS"  
ORIGIN  
Query Match 65.0%; Score 1084.4; DB 9; Length 1651;  
Best Local Similarity 96.7%; Pred. No. 1.5e-307;  
Matches 1133; Conservative 0; Mismatches 1; Indels 38; Gaps 1;  
QY 263 AGGTCATGTTAGCTGGGGAAGTTTGAAGGAAGAGTTTAAAGTTCAGTTTGCAAG 322  
Db 414 AGCTCATGTTGAGCTGGGGAAGTTTGAAGGAAGAGTTTAAAGTTCAGTTTGCAAG 473  
QY 323 ATGACATACAAAAATGAGGAAGACCTTACGATCTTAATCATTTCTTAAGAAAAAG 382  
Db 474 ATGACATACAAAAATGAGGAAGACCTTACGATCTTAATCATTTCTTAAGAAAAAG 533  
QY 383 GATTGACCTTGAACAGGAAAAATGGGAATTGSCACCTACCCCATATAGCTCGT 442  
Db 534 GATTGACCTTGAACAGGAAAAATGGGAATTGSCACCTACCCCATATAGCTCGT 593  
QY 443 GGTCCCGCTACCGGGGAGAGCTGGAGGTTAGGCCATGTGAGCAGATCTGTTTCT 502  
Db 594 GGTCCCGCTACCGGGGAGAGCTGGAGGTTAGGCCATGTGAGCAGATCTGTTTCT 653  
QY 503 TCACCATTAACCGGACCTACCTCATCATCATGACACCAAGCATTCCTTCTCT 556  
Db 654 TCACCATTAACCGGACCTACCTCATCATCATGACACCAAGCATTCCTTCTATGTC 713  
QY 557 -----ATGTACTGACTTTAAACATATAGACTT 584  
Db 714 CCAAGAAATACACACACTTTGGCAGAAATGATGTAAGTCTTTATATAGACTT 773  
QY 585 AACTGTGCTTGGAAAGCCCATCATGACTGGGCTGTTTTCATGAAAGTCCCGAAAAA 644  
Db 774 AACTGTGCTTGGAAAGCCCATCATGACTGGGCTGTTTTCATGAAAGTCCCGAAAAA 833  
QY 645 CAATTATAGCTTTTATTAACCAAGATTTACCTGTTCACTACGCGACGTTTCAG 704  
Db 834 CAATTATAGCTTTTATTAACCAAGATTTACCTGTTTCACTACGCGACGTTTCAG 893  
QY 705 CAGGACATCCCACTGGCACTAATTAACCAATCTTGAGAGCATGGAAGTCTGGAAGT 764  
Db 894 CAGGACATCCCACTGGCACTAATTAACCAATCTTGAGAGCATGGAAGTCTGGAAGT 953  
QY 765 ACTCCGATCTTAATTTCTTTGCAAGTCCAAAAACAGCTTGAAGAAAGATCTCCGCT 824  
Db 954 ACTCCGATCTTAATTTCTTTGCAAGTCCAAAAACAGCTTGAAGAAAGATCTCCGCT 1013  
QY 825 GGTGTATGTACAGTCAAGTGTGACCAACCATCAGACAGGGAAGAGTATGTCAGAGCT 884  
Db 1014 GGTGTATGTACAGTCAAGTGTGACCAACCATCAGACAGGGAAGAGTATGTCAGAGCT 1073  
QY 885 GATGACTTACATGAGGTGATTTCTATGTGTATGTTTACGAAACAAAGACCTCCCTCA 944  
Db 1074 GATGACTTACATGAGGTGATTTCTATGTGTATGTTTACGAAACAAAGACCTCCCTCA 1133  
QY 945 GCAGCTGAAAAATCAGGCTCTATGATGCGGATGCTTTATAGATCATTCACAGTA 1004  
Db 1134 GCAGCTGAAAAATCAGGCTCTATGATGCGGATGCTTTATAGATCATTCACAGTA 1193

QY 1005 TAAATTATCTAGCTTTTGAAGATGCACTTGTGATGACTACATCACTGAGAAAGTTCTG 1064  
DB 1194 TAAGTTATCTAGCTTTTGAAGATGCACTTGTGATGACTACATCACTGAGAAAGTTCTG 1253  
QY 1065 GAGGCACTGAAAGCTGGGGGTAGTCCCTGATATTAACGGAATCCCGACATCAAGACTG 1124  
DB 1254 GAGGCACTGAAAGCTGGGGGTAGTCCCTGATATTAACGGAATCCCGACATCAAGACTG 1313  
QY 1125 GCTTCCAGTAACAAAGTGTATTTCTGTATCAGAAATTTTCTCAACCCCAAGAACTGGC 1184  
DB 1314 GCTTCCAGTAACAAAGTGTATTTCTGTATCAGAAATTTTCTCAACCCCAAGAACTGGC 1273  
QY 1185 AAGTTACATCAGACGCTGGATTTCTGATGACAGATTTGATAGAGCTTATGAGAAATGAA 1244  
DB 1374 AAGTTACATCAGACGCTGGATTTCTGATGACAGATTTGATAGAGCTTATGAGAAATGAA 1433  
QY 1245 GCTGAAGGCTGATCTTAACCAAGCACTTCTGACAGCTCTCAAGGAAACGGAATGGGG 1304  
DB 1434 GCTGAAGGCTGATCTTAACCAAGCACTTCTGACAGCTCTCAAGGAAACGGAATGGGG 1493  
QY 1305 AGTGCAGACGCTCAACCAAGCAATTAATCATGATGATTTGAAGTATGATGACCA 1364  
DB 1494 AGTGCAGACGCTCAACCAAGCAATTAATCATGATGATTTGAAGTATGATGACCA 1553  
QY 1365 GGTGTGGGCTAATATCAGGCTTCAGGAAAGG 1396  
DB 1554 GGTGTGGGCTAATATCAGGCTTCAGGAAAGG 1585

RESULT 9  
AY344579 1389 bp mRNA linear MAM 03-AUG-2003  
LOCUS Bos taurus putative alpha 1,3-fucosyltransferase (Fut10) mRNA,  
DEFINITION complete cds.  
ACCESSION AY344579 GI:33303527  
VERSION AY344579.1 GI:33303527  
KEYWORDS  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 1389)  
AUTHORS Germot,A., Dupuy,F., Lorioi,C., Julien,R. and Matfah,A.  
TITLE Genomic organization and expression profile of putative alpha  
1,3-fucosyltransferase gene, fut10 and fut11, in Bos taurus  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1389)  
AUTHORS Germot,A.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2003) EA 3176 Glycobiologie et Biotechnologie,  
Faculte des Sciences et Techniques, 123, Avenue A. Thomas, Limoges  
87060, France

FEATURES  
source  
1. 1389 Location/Qualifiers  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
1. 1389  
/gene="fuc10"  
1. 1389  
/gene="fuc10"  
/codon\_start=1  
/product="putative alpha 1,3-fucosyltransferase"  
/protein\_id="AA02331.1"  
/db\_xref="GI:33303528"  
/translation="MATVFLITLQVVELGFEKFKRNHLKDGAOMAEPLHLH  
PFNRGLTKRKLTAADSPIMLWSPILGETSGRLGOCADCFPIINVTYFHHK  
TKAFLEFGDFSDSLPLRKAHDMALFHEBSPKNYKLFHGVITLFWTATFSSH  
SHLPLTQYUEGTEVLTLRLHVLPLRSKNHLKSLAPLVYQSCDPPSRDSTVREL  
MTYIEVDSYGECDKRNPLPQSLNSPASMWDGFERILAQYFLLAFENAVACDDYITK  
LMLPLKLVPPVYGSPSIADWLPSNRSAIIVSEPSHRELASYIRALDRDRRYQAY

ORIGIN  
Query Match 60.7%; Score 1013.8; DB 4; Length 1389;  
Best Local Similarity 84.6%; Pred. No. 9.2e-287;  
Matches 1151; Conservative 0; Mismatches 207; Indels 3; Gaps 1;  
STDEFMALVFSD"  
TEMLKGEISNORLITLRRKKGVDVRODNYIDAPEGCWCTKWDNIRLOEKGLP  
KRNQADVTHLSCEPTVTFAPSPVPRRSRLREMIISFQSSKKEAALAMLVDRNQNF

QY 263 AGGTATGCTTGAAGCTGGGGAAGTTTGAAGAAGAGTTTAAAGTTCCAGTTCCAG 322  
DB 32 AGGTATGCTTGAAGCTGGGGAAGTTTGAAGAAGAGTTTAAAGTTTCAATTTTAAAG 91  
QY 323 ATGACATACAAAATGAGAGAACCTTACGATCTTAATTTATTTCTTAAAGAAAG 382  
DB 92 ATGACATACAAAATGAGAGAGCTTCCACCTCCATCCGTTCTTTAATAGAGAG 151  
QY 383 GATTGACCTTCAAGAGAAAGAAATGGAAATTGACAGCTAACCCCATTAAGCTTGGT 442  
DB 152 GATTGACCTTCAAGAGAAAGAAATGGAAATTGACAGCTAACCCCATTAAGCTTGGT 211  
QY 443 GGTCCCGCTGAACGGGGAGACTGGAGCTTAAGCCCAATGAGAGAGAGCTTGTCT 502  
DB 212 GGTCCCGCTGAACGGGGAAACGGGAAGCTGGAGCTTGGAGTGGGAGATGCTTGTCT 271  
QY 503 TCACCATCAACCGGACCTACCTCCATCATCATGACCAAGAAAGATTCCTTCTATGTA 562  
DB 272 TCACCATCAACCGGACCTACCTCCATCATCATGACCAAGAAAGATTCCTTCTATGTA 331  
QY 563 CTGACTTTAATATATATAGCTTACCTTGGCTCGGAAAGCCCATATGATGGCGTGT 622  
DB 332 CTGACTTTAATATATATAGCTTACCTTGGCTCGGAAAGCCCATATGATGGCGTGT 391  
QY 623 TTCATGAGAGTCCCGGAAACAAATTAATAGCTCTTTCATTAACCAAGTATTAAGCT 682  
DB 392 TCACGAGAGAGTCTCCAAAATATATATATAGCTCTTTCATTAACCAAGTATTAAGCT 451  
QY 683 TCACATCACTGCAAGCTTCAAGAGCACTTCCCATCACTATGACCAATTAAGCTTGG 742  
DB 452 TCACATCACTGCAAGCTTCAAGAGCACTTCCCATCACTATGACCAATTAAGCTTGG 511  
QY 743 AGAGATGAGAGTCTGAGTCACTCCGATACCTAGTTCCTTTCAGTCCCAAAACAGC 802  
DB 512 AGAGATGAGAGTCTGAGTCACTCCGATACCTAGTTCCTTTCAGTCCCAAAACAGC 571  
QY 803 TTGAAAAAGACTTGGCTCGGTGTATGATGATGATGATGATGATGATGATGATGATGAT 862  
DB 572 TCCGGAAGAGCTCGCCCGCTGGTGTATGATGATGATGATGATGATGATGATGATGATGAT 631  
QY 863 GGGAGAGCTATGTCGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 922  
DB 632 GGGAGAGCTATGTCGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 601  
QY 923 TACGAAACAAAGACCTCCCTCAGCAGCTGAAATTCAGGCTCTATGATGATGATGATGAT 982  
DB 602 TCCGAAACAAAGACCTCCCTCAGCAGCTGAAATTCAGGCTCTATGATGATGATGATGAT 751  
QY 983 TTTATGATCATTTGACAGAT 1042  
DB 752 TTTATGATCATTTGACAGAT 811  
QY 1043 ACTACATCACTGAGAGAGTTTGAAGGCACTGAAATCGGGGATAGTCCCTGATATATAG 1102  
DB 812 ACTACATCACTGAGAGAGTTTGAAGGCACTGAAATCGGGGATAGTCCCTGATATATAG 871  
QY 1103 GATCCCAAGCATCAAGACTGGCTTCCAGTAAACAAAGTCTATTTCTTGTATCAAGAT 1162  
DB 872 GATCCCAAGCATCAAGACTGGCTTCCAGTAAACAAAGTCTATTTCTTGTATCAAGAT 931  
QY 1163 TTTTCACCCCAAGGAAATGGCAAGTATACATCAAGAGCTGATTTGATGATGATGATGAT 1222  
DB 932 TTTTCACCCCAAGGAGTGGCGAGCTACATCAAGAGCTGATTTGATGATGATGATGATGAT 991

|            |  |   |  |                            |
|------------|--|---|--|----------------------------|
| QY         |  | 1222  | ATGAGGCCTATTGTAAATGGAAAGCTGAAAGGTGAGATCTTACACGAGCACTTTTGACAG     | 1282                       |
| Db         |  | 992   | ACCAGGCTTAATAAATGAAGCTGAAAGGGCGAGATCTGAATTCAGCACAATTTCACACTG     | 10515                      |
| QY         |  | 1283  | CTCTCCGGGAAACGGAAATCGGGAGTGCACAGACGTCACACAGACGAATTAATCATCGATGCAT | 13422                      |
| Db         |  | 1052  | CGCTTAGGGGAAACGGAAGTGGGGGGGTGCAGACGTTTAGACAGGACAACTATCATCGATGGCT | 11111                      |
| QY         |  | 1343  | TTCAGTGTATGCTGTGCACCAAAGTGTGGGCTTAATATCACGGCTTCAGGAAAAGGGCTTAC   | 14022                      |
| Db         |  | 1112  | TTTAGGTATGGTGTGACACCAAGGTGTGGGATTAACATTTAGGCTTCAGGAAAGGGCTTAC    | 11712                      |
| QY         |  | 1403  | CACCCAAAAGATGGGAGGACAGAGAATACCCACCTGAGATTCGCCAGAGCCACAGTGTTCG    | 14622                      |
| Db         |  | 1172  | CACCCAAAAGATGGGAGGACAGACGTTTACCCACCTGAGCTTCCAGAGCTTACAGTGTTCG    | 12313                      |
| QY         |  | 1463  | CTTTCTCACCACTCCGAGACTCCACTTTGAGCTCTTTGCGAGAGATGTGGATTTCCAGCT     | 15222                      |
| Db         |  | 1232  | CTTTCTCACCC---CCCTGTGCTCCAGCTCCGGCCCTCTTTGCGAGAGATGTGGATTCACCACT | 12882                      |
| QY         |  | 1523  | TTTCAATATCCCAAGAAAGAACGCCAGGACATCAAGTGGCTGTGATPAGAAATCAAACT      | 15822                      |
| Db         |  | 1289  | TTCAGAAATCCCAAGAAAGAACGCCAGGACATCAAGTGGCTGTGATPAGAAATCAAACT      | 13482                      |
| QY         |  | 1583  | TTTTATCTCAAGAGTTTGGGGCCCTAGTATTCAGAGACTGA                        | 1623                       |
| Db         |  | 1349  | TTTCAACTCAAGAGTTTGGGGCCCTAGTATTCAGAGACTGA                        | 1389                       |
| RESULT 10  |  |   |  |                            |
| LOCUS      |  | AC067838  | 104120 bp  | DNA linear PRI 21-MAR-2003 |
| DEFINITION |  | Homo sapiens chromosome 8, clone Rpl1-722E23, complete sequence.  |  |                            |
| VERSION    |  | AC067838  |  |                            |
| ACCESSION  |  | AC067838.18   | GI:29135646  |                            |
| KEYWORDS   |  | HTG.  |  |                            |
| SOURCE     |  | Homo sapiens (human)  |  |                            |
| ORGANISM   |  | Homo sapiens  |  |                            |
| REFERENCE  |  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   |  |                            |
| AUTHORS    |  | Mammalia; Eutheraia; Primates; Catarrhini; Homnidae; Homo.  |  |                            |
| JOURNAL    |  | 1 (bases 1 to 104120)   |  |                            |
| REFERENCE  |  | Birren,B., Nusbaum,C. and Lander,E.<br>Homo sapiens chromosome 8, clone Rpl1-722E23<br>Unpublished<br>2 (bases 1 to 104120)   |  |                            |
| AUTHORS    |  | Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,<br>Anderson,S., Balgwin,J., Barna,N., Bastien,V., Beda,F.,<br>Boguslavsky,I., Bouckgalter,B., Brown,A., Burkett,G.,<br>Campiano,A., Castle,A., Choppe,Y., Colangelo,M., Collins,S.,<br>Collymore,A., Cooke,P., Dekrelano,K., Dewar,K., Diaz,J.S.,<br>Dodges,S., Domino,M., Doyle,M., Ferreira,P., Fitchugh,W., Gage,D.,<br>Galagan,J., Gardyna,S., Glnde,S., Guyette,M., Graham,L.,<br>Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,<br>Howard-Jones,R., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,<br>Klein,T.J., LaRoque,K., Lamazares,R., Landers,C., Lehotzky,J.,<br>Levine,R., Lieu,C., Liu,C., Locke,K., MacDonald,T., Margulis,N.,<br>McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McNetters,R.,<br>Melchior,J., Menesius,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,<br>Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,<br>O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Plante,N.,<br>Pisanti,C., Pollara-V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,<br>Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,<br>Stange-Rothmann,N., Stojanovic,N., Subramanian,A., Talmas,J.,<br>Testafaye,S., Theodore,J., Tittelli,A., Travers,M., Trigilio,J.,<br>Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye.W.J.,<br>Young,G., Zainoun,J., Zimmer,A. and Zody,M. |  |                            |
| TITLE      |  | Direct Submission   |  |                            |
| JOURNAL    |  | Submitted (27-Apr-2000) Whitehead Institute/MIT Center for Genome   |  |                            |
| REFERENCE  |  | Research, 320 Charles Street, Cambridge, MA 02141, USA  |  |                            |
| AUTHORS    |  | 3 (bases 1 to 104120)<br>Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,<br>Barna,N., Bastien,V., Bloom,T., Boguslavsky,I., Bouckgalter,B.,<br>Camarate,J., Chang,T., Chazaro,B., Choppe,Y., Collymore,A.,   |  |                            |

COOK, A., COOKE, P., DEARELLANO, K., DEWAR, K., DIAZ, J.S., DODGE, S., FAIRO, S., FERREIRA, P., FITZGERALD, M., GAGE, D., GALAGAN, J., GARDYNA, S., GORD, S., GRATHAM, L., GRAND-PIERRE, N., HAGOS, B., HORTON, L., HULME, W., ILIEV, I., JOHNSON, R., JONES, C., KAMET, A., KARATAS, A., KELLS, C., LANDERS, T., LEVINE, R., LINDBLAD-TOH, K., LIU, G., MACLEAN, C., MACDONALD, P., MAJOR, J., MATHEWS, C., MCCARTHY, M., MELDIRIM, J., MENEUS, L., MHOVA, T., MLENGA, V., MURPHY, T., NAYLOR, J., NGUYEN, C., NICOL, R., NORBU, C., NORMAN, C.H., O'CONNOR, I., O'DONNELL, P., O'NEILL, D., OLIVER, J., PETERSON, K., PHUNKHANG, P., PIERRE, N., RAYMOND, C., RETTA, R., RISE, C., ROGOV, P., ROMAN, J., ROY, A., SCHAUER, S., SCHUPBACK, R., SEAMAN, S., SEVERY, P., SMITH, C., SPENCER, B., STANGE-THOMANN, N., STOJANOVIC, N., TALAMAS, J., TESTAYE, S., THEODORE, J., TOPHAM, K., TRAVERS, M., VASSILIEV, H., VIEL, R., VO, A., WILSON, B., WU, X., WYMAN, D., YOUNG, G., ZAINOUN, J., ZEMBEK, L., ZIMMER, A. and ZOXY, M.

**TITLE**  
Direct Submission

**JOURNAL**  
Submitted (10-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

**REFERENCE**  
4 (bases 1 to 104120)

**AUTHORS**  
Birren, B., Nussbaum, C., Landers, E., Ali, A., Allen, N., Anderson, S., Banna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalder, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collumore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fair, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Gratham, L., Grand-pierre, N., Hafez, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Menus, L., Mhova, T., Mlenga, V., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zoxy, M.

**TITLE**  
Direct Submission

**JOURNAL**  
Submitted (03-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

**REFERENCE**  
5 (bases 1 to 104120)

**AUTHORS**  
Birren, B., Nussbaum, C., Landers, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Banna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalder, B., Camarata, J., Chang, J., Choepel, Y., Collumore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Eriksson, J., Fair, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gratham, L., Grand-pierre, N., Hafez, N., Hascopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Menus, L., Mhova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, J., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Scubbs, R., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zoxy, M.

**TITLE**  
Direct Submission

**JOURNAL**  
Submitted (12-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

**COMMENT**  
On Mar 21, 2003 this sequence version replaced g1:24496795. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information  
Center project name: L9405  
Center clone name: 722\_E23  
-----

Only the first 104.1 kilobases of the middle contig is being submitted.

This clone is completely overlapped by number AC09114 [WICR project L13069] and AC104027 [WICR project L21993].

## FEATURES

Location/Qualifiers  
1..104120  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-722E23"  
9..142  
/clone\_lib="RPC1-11 Human Male BAC"  
repeat\_region  
/rpt\_family="(TA)n"  
complement(123..378)  
/rpt\_family="L2"  
repeat\_region  
complement(865..923)  
/rpt\_family="MLTIF"  
repeat\_region  
complement(984..1374)  
/rpt\_family="MLTIF"  
repeat\_region  
complement(1401..1576)  
/rpt\_family="MLTIF-int"  
repeat\_region  
complement(1940..2100)  
/rpt\_family="MLTIF1"  
2161..2427  
/rpt\_family="L2"  
repeat\_region  
complement(4278..4692)  
/rpt\_family="MLTIF1"  
4818..4937  
/rpt\_family="L2"  
repeat\_region  
4941..5075  
/rpt\_family="MIR"  
repeat\_region  
6708..6797  
/rpt\_family="L2"  
repeat\_region  
complement(6892..6955)  
/rpt\_family="L2"  
repeat\_region  
complement(7144..7447)  
/rpt\_family="Aluuc"  
7745..7854  
/rpt\_family="L2"  
repeat\_region  
complement(7907..7982)  
/rpt\_family="MIR"  
8737..8773  
/rpt\_family="(CAAAA)n"  
repeat\_region  
complement(9080..9150)  
/rpt\_family="MLTIF"  
repeat\_region  
complement(10454..10537)  
/rpt\_family="L2"  
repeat\_region  
10538..10843  
/rpt\_family="AluSg"  
repeat\_region  
complement(10844..10894)  
/rpt\_family="L2"  
10971..11034  
/rpt\_family="MIR"  
11247..11379  
/rpt\_family="L2"  
11391..11425  
/rpt\_family="AT-rich"  
11434..11721  
/rpt\_family="AluSg"  
11916..12037  
/rpt\_family="Aluuc"  
repeat\_region  
complement(12038..12350)  
/rpt\_family="AluY"  
12351..12539  
/rpt\_family="Aluuc"  
12540..12559  
repeat\_region

repeat\_region  
12578..12889  
/rpt\_family="AluSx"  
repeat\_region  
12903..12923  
/rpt\_family="(GGGA)n"  
repeat\_region  
complement(13576..13750)  
/rpt\_family="L2"  
repeat\_region  
complement(13954..14317)  
/rpt\_family="L2"  
repeat\_region  
complement(14533..14599)  
/rpt\_family="MER5A"  
repeat\_region  
complement(14600..14891)  
Query Match  
Best Local Similarity 99.2%; Pred. No. 4,9e-235;  
Matches 843; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
547 TTCTCTTCTATGCTACTGACTTAAACATAGATAGCTTACCTGTGCTGGAAAGCCCAT 606  
70187 TTCTCTTCTCTAGGACTGACTTAAACATAGATAGCTTACCTGTGCTGGAAAGCCCAT 70128  
607 CATGACTGGGCTGTTTCAAGAGAGTCCCGGAAACAAATATAGCTCTTCAATAA 666  
70127 CATGACTGGGCTGTTTCAAGAGAGTCCCGGAAACAAATATAGCTCTTCAATAA 70068  
667 CCAGTATTAACCTGTTCAACTACACTCCACGCTTCAGCAGGCAATCCACTTGCACATA 726  
70067 CCAGTATTAACCTGTTCAACTACACTCCACGCTTCAGCAGGCAATCCACTTGCACATA 70008  
727 ACTACCAATCTTGGAGAGATGAACTCTGAAGTCACTCCGATACCTAGTCTCTTGG 786  
70007 ACTACCAATCTTGGAGAGATGAACTCTGAAGTCACTCCGATACCTAGTCTCTTGG 69948  
787 CAGTCCAAAACAAGCTTAGAAAAGACTGCTCCGCTGCTGATGATACAGTCAAGTGT 846  
69947 CAGTCCAAAACAAGCTTAGAAAAGACTGCTCCGCTGCTGATGATACAGTCAAGTGT 69888  
847 GACCCACCATAGACAGGAGACAGTATGTTGCGAGCTGATGACTTACATCGAGTCTGAT 906  
69887 GACCCACCATAGACAGGAGACAGTATGTTGCGAGCTGATGACTTACATCGAGTCTGAT 69828  
907 TCCATGCTGATGATGTTACGAAACAAGACCTCCCTCAGCAGCTGAAAATCCAGCTCT 966  
69827 TCCATGCTGATGATGTTACGAAACAAGACCTCCCTCAGCAGCTGAAAATCCAGCTCT 69768  
967 ATGATGCCGATGCTTTTATAGATCATTGACAGTAAAGTTTATCTAGCTTTTGA 1026  
69767 ATGATGCCGATGCTTTTATAGATCATTGACAGTAAAGTTTATCTAGCTTTTGA 69708  
1027 AATGAGTTTGTGATGACTACATCAGTAAGTTCGAGGCGCACTGAAATCGGGGATA 1086  
69707 AATGAGTTTGTGATGACTACATCAGTAAGTTCGAGGCGCACTGAAATCGGGGATA 69648  
1087 GTCCCTGATATTAACGATCCCGCAGCATCAGACTGCTGCTTCCAAATGAAAGTGT 1146  
69647 GTCCCTGATATTAACGATCCCGCAGCATCAGACTGCTGCTTCCAAATGAAAGTGT 69588  
1147 ATCTTGTATAGAAATTTTTCACCCCAAGGAAGTTCAGAGTTACATCAGACATCTGAT 1206  
69587 ATCTTGTATAGAAATTTTTCACCCCAAGGAAGTTCAGAGTTACATCAGACATCTGAT 69528  
1207 TCTGATGACAGATGTTAGAGGCTTATGTAAGTGAAGCTGAAGGTGAGATCTCTAAC 1266  
69527 TCTGATGACAGATGTTAGAGGCTTATGTAAGTGAAGCTGAAGGTGAGATCTCTAAC 69468  
1267 CAGCAGCTTCTGACAGCTTCTAGGAAACGAAATGGGAGTGCAGACGTCACCGAGAC 1326  
69467 CAGCAGCTTCTGACAGCTTCTAGGAAACGAAATGGGAGTGCAGACGTCACCGAGAC 69408  
1327 AATTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1386  
69407 AATTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 69348



|            |   |                  |       |  |  |  |  |
|------------|---|------------------|-------|--|--|--|--|
| OY         |   | 1387 CAGGAAAAGC  | 1396  |  |  |  |  |
| Db         |   | 69347 CAGGAAAAGC | 69338 |  |  |  |  |
| RESULT 11  |   |                  |       |  |  |  |  |
| LOCUS      | AC091144  |                  |       |  |  |  |  |
| DEFINITION | Homo sapiens chromosome 8, clone RP11-359B20, complete sequence.  |                  |       |  |  |  |  |
| ACCESSION  | AC091144  |                  |       |  |  |  |  |
| VERSION    | AC091144.12   |                  |       |  |  |  |  |
| KEYWORDS   | GI:24580467   |                  |       |  |  |  |  |
| ORGANISM   | HTG.  |                  |       |  |  |  |  |
| SOURCE     | Homo sapiens (human)  |                  |       |  |  |  |  |
| REFERENCE  | Homo sapiens  |                  |       |  |  |  |  |
| AUTHORS    | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.   |                  |       |  |  |  |  |
| TITLE      | 1 (bases 1 to 185872)   |                  |       |  |  |  |  |
| JOURNAL    | Homo sapiens chromosome 8, clone RPl1-359B20  |                  |       |  |  |  |  |
| REFERENCE  | Unpublished   |                  |       |  |  |  |  |
| AUTHORS    | 2 (bases 1 to 185872)   |                  |       |  |  |  |  |
| -          | Bairren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Brown,A., Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dekrelano,K., Dewark,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,S., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-pierre,N., Hagos,B., Haeftord,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., Lapocque,K., Lamaras,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Marguis,N., Matthews,C., McCarthy,M., McEwan,P., McLernan,K., McPheters,R., Meldrum,D., Menus,L., Mihowa,T., Mengra,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunphang,P., Piere,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schaner,S., Schnuppach,R., Seaman,S., Severi,P., Sougnere,C., Spencer,B., Strange-Thomann,N., Scjojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.Y., Young,G., Zainoun,J., Zemck,L., Zimmer,A. and Zody.M. |                  |       |  |  |  |  |
| TITLE      | Direct Submission   |                  |       |  |  |  |  |
| JOURNAL    | Submitted (01-APR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  |                  |       |  |  |  |  |
| REFERENCE  | 3 (bases 1 to 185872)   |                  |       |  |  |  |  |
| AUTHORS    | Bairren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barrna,N., Bastien,V., Bloom,T., Bogslaxkiv,L., Bolkhalter,B., Camarata,J., Chang,J., Charazo,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArizcano,K., Dewark,K., Diaz,J.S., Dodge,S., Ferro,S., Ferreira,P., Fritzcerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hatzek,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toch,K., Liu,G., Maclean,C., MacDonald,P., Major,T., Matthews,C., McCarthy,M., Meldrim,J., Menus,L., Mihowe,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norma,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunphang,P., Piere,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Santos,R., Schaner,S., Schnuppach,R., Seaman,S., Severi,P., Sougnere,C., Spencer,B., Strange-Thomann,N., Scjojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.Y., Young,G., Zainoun,J., Zemck,L., Zimmer,A. and Zody.M.  |                  |       |  |  |  |  |
| TITLE      | Direct Submission   |                  |       |  |  |  |  |
| JOURNAL    | Submitted (25-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  |                  |       |  |  |  |  |
| REFERENCE  | 4 (bases 1 to 185872)   |                  |       |  |  |  |  |
| AUTHORS    | Bairren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barrna,N., Bastien,V., Bloom,T., Bogslaxkiv,L., Bolkhalter,B., Camarata,J., Chang,J., Charazo,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArizcano,K., Dewark,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fritzcerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hatzek,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toch,K., Liu,G., Maclean,C., MacDonald,P., Major,T., Matthews,C., McCarthy,M., Meldrim,J., Menus,L., Mihowe,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norma,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunphang,P., Piere,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Santos,R., Schaner,S., Schnuppach,R., Seaman,S., Severi,P., Sougnere,C., Spencer,B., Strange-Thomann,N., Scjojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.Y., Young,G., Zainoun,J., Zemck,L., Zimmer,A. and Zody.M.   |                  |       |  |  |  |  |

```

TITLE
JOURNAL
COMMENT
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: I13069
Center clone name: 355_B_20
-----
Location/Qualifiers
1. 185872
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="8"
   /map="8"
   /clone="RP11-359B20"
   /clone.lib="RPCI-11 Human Male BAC"
repeat_region complement(1..273)
               /rpt_family="MER50"
repeat_region complement(251..503)
               /rpt_family="MER50"
repeat_region 521..549
               /rpt_family="(A)n"
repeat_region 561..812
               /rpt_family="AlusX"
repeat_region complement(815..918)
               /rpt_family="(L2)"
repeat_region 924..1036
               /rpt_family="FLAM_C"
repeat_region complement(1132..1688)
               /rpt_family="MLT2B2"
repeat_region complement(1694..1991)
               /rpt_family="Aluo"
repeat_region complement(2063..2368)
               /rpt_family="AlusX"
repeat_region 2398..2479
               /rpt_family="(LMC4a)"
repeat_region 2480..2740
               /rpt_family="AlusC"
repeat_region 2741..3183
               /rpt_family="(LMC4a)"
repeat_region 3202..3323
               /rpt_family="FLAM_C"
repeat_region 3324..3352
               /rpt_family="(7AAA)n"
repeat_region 3398..3636
               /rpt_family="(LMCS)"
repeat_region 3775..3857
               /rpt_family="(LMCS)"
repeat_region 3862..4125
               /rpt_family="AlusX"
repeat_region 4127..4439

```





Db 1144 ATTACATACAGAGAAATTCTCGAGACCACTGAAACTGGGGGTCGTGCCCTGTCTTACG 1203

Qy 1103 GATCCCCCAGCATCAGACTGGCTCTTCAAGTAACAAAGTCTATTCTTGTATCAGAAAT 1162

Db 1204 GATCCCCCAGCATCAGACTGGCTCTTCAAGTAACAAAGTCTATTCTTGTATCAGAAAT 1263

Qy 1163 TTCTCAGCCCGAGGAACTGGCAAGTATCATCATCAGAGACTGATCTGATGACATTTGT 1222

Db 1264 TCTTCAACCCCAAGAGACTGGCAAGTCTCATTTAGAGACTGATATATGACATGGGTTGT 1323

Qy 1223 ATGAGGCTTATGTAGAAATGAGCTGAAGGCTGATCTCTTAACCAAGCACTTCTGACAG 1282

Db 1224 ATGAGACTTATGTAGAAATGAGCTGAAGGCTGATCTCTTAACCAAGCACTTCTGACAG 1383

Qy 1283 CTCCTCAGGAAACGGAATGGGGAGTGCAGACGTCAACAGCAATTTACATGATGAT 1342

Db 1384 CTCCTCAGGAAACGGAATGGGGAGTGCAGACGTCAACAGCAATTTACATGATGATGAT 1443

Qy 1343 TTGAGTGTATGTGTGTGACCAAGGTGTGGCTTAATATCAGGCTTACAGAAAG 1396

Db 1444 TTGAGTGTATGTGTGTGACCAAGGTGTGGCTTAATATCAGGCTTACAGAAAG 1497

RESULT 13

BC062113 2978 bp mRNA linear ROD 11-DEC-2003

LOCUS MMS musculus fucosyltransferase 10, mRNA (cDNA clone MGC:69807

DEFINITION IMAGE:6816067), complete cds.

ACCESSION BC062113

VERSION BC062113.1 GI:38328410

KEYWORDS MGC.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 2978)

AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heien, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Shapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., USCIN, T.B., Tosinylki, S., Gernluch, P., Frange, C., Rana, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Muliany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huliy, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Boultard, G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skaleka, U., Smalins, D.E., Scherch, A., Schein, J.F., Jones, S.J. and Marra, M.A.

TITLE human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

PUBMED 12477932

REFERENCE 2 (bases 1 to 2978)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-2003) National Institutes of Health, Mammalian Gene Collection (MGC) Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (IHLN)

DNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada

info@bcgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdel, Yaron Butterfield, Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carlie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Jissa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyerdun, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/IHLN at: <http://image.llnl.gov>

Series: IRAX Plate: 132 Row: h Column: 16

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

source location/Qualifiers

1..2978

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="MGC:69807 IMAGE:6816067"

/tissue\_type="Brain, mouse, 13.5, 14.5, 16.5, 17.5 dpc"

/clone\_id="NIH BMAP\_FMO"

/lab\_host="DH10B"

/note="vector: pYX-ASC"

1..2978

/gene="Fuc10"

/note="synonym: MGC40819"

/db\_xref="locusID:171167"

/db\_xref="MGI:2384748"

387..1697

/codon\_start=1

/product="fucosyltransferase 10"

/protein\_id="AAH62113.1"

/db\_xref="GI:38328411"

/db\_xref="locusID:171167"

/translation="MVRFORKSLASCLCTATVPLVTLQVVELGKPERKKLXDSN VODRIVDEGPKLEDEPREKALAGRKNVAGSYPIVLMSPITGTEGRGQCGA DACEFTTNETQHPMRAPLFTGTDINDISLPEAFHMDLFEESRKNYKLFH KPVITREHRTREKSHLPLTTOYLEGVDVSKSLIVPLQKNNLRQKLAPLVYQ SDQCPDSRYSVLEAVYEVDSYGSCLDNRPLPOOLKNPAMDADAEVRAVQKF ILAEWVCDYIIEKFWRLKGVVYVYGSPTIADWPSNRSATVGFSPRLA SFIRILDYDGLVETVYEWRLKGISNORLLTLNREMGVQDINDNYIDSEWCW RRVWANSRLQGVSEMRKSGMHGSLCVLFLFLLMWLPATGLYS"

633..1574

/note="Glyco\_transf.10; Region: Glycosyltransferase family 10 (fucosyltransferase). This family of fucosyltransferases are the enzymes transferring fucose from GDP-Fucose to GlcNAc in an alpha1,3 linkage. This family is known as glycosyltransferase family 10"

/db\_xref="CDP:pfam00852"

ORIGIN

Query Match 47.9%; Score 799.6; DB 10; Length 2978; Best Local Similarity 81.6%; Pred. No. 1,2e-223; Matches 925; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

Qy 263 AGTCATGCTGAGCTGGAGAGTTTAAAGAGAGAGTTTAAAGTCCAGTTTGCAG 322

Db 466 AGTAGTGTGTGAGCTGGAGAGTTTAAAGAGAGAGTTTAAAGTCCAGTTTGCAG 525

Qy 323 ATGACATCAAAAATGAGAGAGACCTTACGATCTTAATTATTTCTTAAAGAAAG 382

Db 526 ACGGACATAGAGATGTGAGAGAGAGCTTAAACCTGGAGCCGTTTCTGAAAAGAG 585

Qy 383 GATTGACCTTCAACAGGAAAGAAATGAGAGAGCTTCAAGCTCCCATATGCTGTGT 442

Db 586 CACTGGCTTGTGCTGGAGAGCAAGAGTGTGATCTTACCCATCGTCTGTGT 645

Qy 443 GGTCCCGCTGACGGGGAGACTGGAGAGTTTAAAGTCCATGTGAGAGAGATGCTTGT 502

```

Db      646 GGTCCCACTGAGTGGAGAGACTGGAGGCTTGGCCAGTGTGGGGAGATGCTTCTTCT 705
QY      503 TTACCATACACCGGACCTTACCTCATCATGACCAAGATTCCTTCTTATGGTA 562
Db      706 TTACCATACACCGGACCTTCCAGATCATCGAGACAGGCGCTTCTTATTAAGGTA 765
QY      563 CTGACTTTAAGATAGATAGTACTTACCTCTGCTCGGAAGGCCATCATGATCGGGCTGTT 622
Db      766 CTGACTTTAAGATAGAGCTTACCTTGTGCTCGGAAGGCCATCATGATCGGGCGCTTT 825
QY      623 TTATGAAAGTCCCGGAAAAAACAATTATAAGCTCTTTTCAATTAACCGATATACCTTGT 682
Db      826 TCCATGAAGAGCGCCCAAAAAACAATTATAAGCTCTTCAATTAACCGATATACCTTAT 885
QY      683 TCAACTACACTGCGCAGTTCAGACAGGCGATTCGCCACTTGCCACTACTACCAATACCTGG 742
Db      886 TCAACCATACCGCCCATTTCCAGACAGGATTCGCCACTTGCCACTACTACCAATACCTGG 945
QY      743 AGAGCATTTGAAGTCTGAAGTCACTCCGATACCTAGTTCCTTTGAGCTCCAAAAACAAGC 802
Db      946 AAGCGTGGAGCTGCTGAAGTCACTCAGATACCTAGTTCCTTTACAGGCTAAGAACACAC 1005
QY      803 TTAGAAAAAGCTGCTGCGGTGATATGATACAGTCAAGTGCACCGACATCCAGACA 862
Db      1006 TTAGAAAAAGTGGCGCCCATTTGATATGTTCACTGAGTGCATCCATCCATCCAGACA 1065
QY      863 GGGACAGCTATGTTGCGCAGCTGATGACTTACATGAGGTCGATTCCTATGATGATGTT 922
Db      1066 GGGACAGCTATGTTGCGGAGGCTGATGCGTACATTTGAAGTTCCTATGCGAGATGTT 1125
QY      923 TTGGAACAAAGACTCTCCTCAGACGCTGAAAAATCCAGCCTTATGGAATGCCATGCT 982
Db      1126 TTACAAACAGAAATTTCCCAAGCAGCTGAAAAACCCCGCTCCATGGAATGCCATGCT 1185
QY      983 TTTATAGATCATTTGCAAGTATAAATTATCTTATCTTATGAGATGAGCAGTTTGTATG 1042
Db      1186 TCTACAGATCATTTGCCAGATTAAGTTCACTCGGCTTTGAAATGCCGCTGTGACG 1245
QY      1043 ACTACATCACTGAGAACTTTGAGGCGCATGAAATCGGGGGTATGCCCTGTATATTAG 1102
Db      1246 ATTACATCACTGAGAACTTTGAGGCGCATGAAATCGGGGGTATGCCCTGTATATTAG 1205
QY      1103 GATCCCCCAGCATCAAGCTGAGCTTCCAGATTAACAAAGTCTATTCTGTATCAGAT 1162
Db      1306 GATCCCCCAGCATCAAGCTGAGCTTCCAGATTAACAAAGTCTATTCTGTATCAGAT 1365
QY      1163 TTTCTACCCCCAGGAGCTGCGCAAGTTACATCAAGCACTGATTCGTATGACAGATTGT 1222
Db      1366 TCTCTACCCCCAGGAGCTGCGCAAGTTACATCAAGCACTGATTCGTATGACAGATTGT 1425
QY      1223 ATGAGGCTTATTAATGAGAGCTGAGAGGAGTGAATCTTAACCAAGCACTTCTGACAG 1282
Db      1426 ATGAGGCTTATTAATGAGAGCTGAGAGGAGTGAATCTTAACCAAGCACTTCTGACAG 1485
QY      1283 CTCTCAGGAAACGAAATGAGAGAGTGAAGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAG 1342
Db      1486 CTCTCAGTAAAGGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAG 1545
QY      1343 TTGAGTGTATGTTGACCAAGGTTGAGGCTTATATAGCTTCAAGAAAAAG 1396
Db      1546 TTGAGTGTATGTTGACCAAGGTTGAGGCTTATATAGCTTCAAGAAAAAG 1599

```

```

RESULT 14
GGA535692      1657 bp      mRNA      linear      VRT 23-DEC-2002
LOCUS      GGA535692
DEFINITION      Gallus gallus mRNA for putative alpha3-fucosyltransferase (FUT10
gene).
ACCESSION      AJ535692
VERSION      AJ535692.1 GI:27368920
KEYWORDS      alpha3-fucosyltransferase; FUT10 gene.
SOURCE      Gallus gallus (chicken)
ORGANISM      Gallus gallus

```

```

REFERENCE
1 Candelier, J.J., Martinez-Duncker, I., Oriol, R., and Mollitone, R.
  Cloning expression and genomic organization of a new human
  alpha3-fucosyltransferase (FUT10)
  Unpublished
  2 (bases 1 to 1657)
  Oriol, R.
  Direct Submission
  Submitted (19-DEC-2002) Oriol, R., Glycobiology, INSERM U504, 16 Av.
  Paul Valliant-Couturier, 94807 Villejuif, FRANCE
  Related sequence: AJ512465.
  Location/Qualifiers
    1..1657
      /organism="Gallus gallus"
      /mol_type="mRNA"
      /db_xref="taxon:9031"
    1..1657
      /gene="FUT10"
    210..1529
      /gene="FUT10"
      /function="adds fucose in alpha1-3 linkage to GlcNAc"
      /codon_start=1
      /product="putative alpha3-fucosyltransferase"
      /protein_id="CAD59689.1"
      /db_xref="GI:27368921"
      /db_xref="GOA:O8AMB5"
      /db_xref="SPTRMBL:O8AMB5"
      /translation="MYMRERKRLMASGICGPAFFPLVLTLOVITELSENKPAVSS
      LHSGLPLDERHNSQLKNELYNFRPEPTDHPILKMLSPRTGEGSGGCEHVC
      FETINKYVSHOMTRAPLFYGDPSIDSLPDKHDMALRHEEPKNNYKLFHEPA
      ITLEFNHATFSRSHSLPITTOYLESLEVSLESHMTLPVQKNSLRRLAPLVQSDC
      NAFCDRDSYVRELMCHIEVDSYGCLEHNDLPQHLNPSAMDNDGNEFKYLAQYKFLI
      FENACEDSYIREKLMLGVVYVYFGSPSIIIDWLPSKSAIVLVSFHPRLAYI
      KTLPNQOEYEAYLEMKLKGDISNPRLITMKRKMGVQDVTDNYIDTFECWVQNRV
      MENTRERKEMLPQRMSAQVYNHNLNCPRLRSSGSLPIPARAPSKRG"
    1632..1637
      /gene="FUT10"
    1653
      /gene="FUT10"
  polyA_signal
  polyA_site
  polyA_site
  ORIGIN
  Query Match. 39.6%; Score 661; DB 5; Length 1657;
  Best Local Similarity 68.6%; Pred. No. 7.2e-103;
  Matches 946; Conservative 0; Mismatches 420; Indels 13; Gaps 2;
  263 AGGTGATGTTGAGCTGGGAAAGTTGAAAGAGAGTTTAAAGTTCCAGTTTGCAG 322
  Db      269 AGGTATACATGAACTGGGAAATTCGAGAAATAGGACACAGCAGTCTCCAGTTTACACA 348
  QY      323 ATGACATACAAAAATGAGAGAGACACTTACGATCTTAATTCATTTCTTAAGAAAAAG 382
  Db      349 GTGGGCCACTGAAGCTGATAGAGACATG-----TTGGCAGCTTAAGAAAGATG 399
  QY      383 GATTGACCTTCAACAGAGAAAGAAATGAGATGACAGTACCCCATTAATGCTGTG 442
  Db      400 AACTTTATAGCAATTCAGACAGAGCTGATACAGATCACTACCTACCTGTGCT 459
  QY      443 GGTCCCGCTGACGGGGAGACTGGAGAGTTAGGCCAATGTGAGAGCATGCTTTTCT 502
  Db      460 GGTCCTCCCTGACTGAGAGAGACAGGAGAGTGGGGCAGTGTGGAGAGATGTTGTTCT 519
  QY      503 TCACCATCAACCGGACCTTACCTCATCATGACCAAGCAATTCCTCTATAGTA 562
  Db      520 TTACTATTAAACAAGCGTACACAGACAGTCAAGATGACAAAGCTTTTGTGTTATAGTA 579
  QY      563 CTGACTTTAAGATAGATGCTTACCTCTGCTCGAAAGCCCATCATGACTGGGCTGTT 622
  Db      580 CTGACTTTAAGATAGACGCTTACCCCTCTCGAAAAACCATCATGATGGGCCCTTT 639
  QY      623 TTCATGAAGAGTCCCGGAAAAACAATTATAAGCTCTTCAATAACCAAGTATTAACCTTGT 682

```

```

Db      640 TCATGAGGATACACCAAAAAACACTACAACTTTTCCATGACACAGATACACTTAT 699
Qy      683 TCAACTACATGCGACAGTTCAGAGGACATTCCTCCTTGGCCATTAACCAACTACTGG 742
Db      700 TCAACCACTGCAACTTTCAGCGCCATTCCTACCTACCGTGCACACTCAGTACCTTG 759
Qy      743 AGAGCATGGAAGTCTGCAAGTCACTCCGATACCTAGTCTTTCGACATCCAAAAACAAGC 802
Db      760 AGAGCATGAGAGTCTCAGGTCCTTAGACATAATGATCCAGTGCAGATGAGAACAACC 819
Qy      803 TTGAAAAAAGTGTGCTCCGCTGGTATGATGACAGACAGCTGTGACCCACATCAGACA 862
Db      820 TGAGGAAGCGGCTTGACCACTTGCTATGATGCAAGTGCATGCAATGCTCTTCTGACC 879
Qy      863 GGGACAGCTATGTTCCGAGCTGATGACTTAACATGAGTGCATTCCTATGCTGATGTT 922
Db      880 GGGACAGCTATGATGCTGAGTGTGATGTCACATTTGAAGTACCTTTACGAGATGTC 939
Qy      923 TAGAAAAAAGAGCTCCCTCAGACAGCTGAAAAATCCAGCCCTCTATGATGCCATGCT 982
Db      940 TGCAATACAGAGACTTCTTCACACTTCAGAAATTCATCTGCCATGATATGAGAACT 999
Qy      983 TTATAGGATCATTGACAGATTAAGTTTATCTTCTAGCTTTTGAATGACATTTGATG 1042
Db      1000 TCTATAAATACTGGCAGACAGTACAAAGTTCATTCTTGCTTTGAAAAATGCTATCTGTAG 1059
Qy      1043 ACTACATCATGGAAGTCTTGAGAGCCACTGAAATCTGGGGTAGTCCCTGTATATACG 1102
Db      1060 ATTATATCATGAAAAACTCTGGCGCTCTATGCTGAGATGTTACTTGTGACTTGG 1119
Qy      1103 GATCCCGACATCACAGACTGGCTCCAGTAAAGAAAGTCTATTTCTTATCAGAT 1162
Db      1120 GTTCTCCGACATTTACGACTGGCTCCAGACAAAGATGCAATCTGTATCTAGTT 1179
Qy      1163 TTTCTCACCCAGGGAAGTGGCAATTAATACAGCACTGGATTTCTGATACAGATTTG 1222
Db      1180 TTTTACACCTCCAGAGCTGGCCCGCTATATCAAAAACCTGGATCAAAATACACAGAGT 1239
Qy      1223 ATGAGGCTATGTAGAATGGAAGTGAAGTGAAGTCTTACACAGGCACTTCTGACAG 1282
Db      1240 ATGAGGCTATGCTGGAATGGAATGGAAGTGAAGTGAAGTCAATTCAGATGAGCTGTACG 1299
Qy      1283 CTCTCAGGGAACGGAATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1342
Db      1300 CAATGAAGGAACGGAATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1359
Qy      1343 TTGAGTGTATGCTGTCACCAAGTGTGGGCTAATATCAGGCTTCAGAAAAAGGCTTAC 1402
Db      1360 TTGAGTGTATGCTGTCACCAAGTGTGGGCTAATATCAGGCTTCAGAAAAAGGCTTAC 1419
Qy      1403 CACCAAAAAGATGGAAGGAGGAAGATACCACTGATGCTGCTGCTGCTGCTGCTGCTGCT 1462
Db      1420 TGCCCGAGAGTGAAGTGTCTGAGTTAACCATCTGACCTGCAAACTGAGGCTTCTG 1479
Qy      1463 CTCTTTCACACTCCGGAATCACTTTGAGTCTTTTGGGAGATGTTGATTTCCAGCT 1522
Db      1480 GTTCTCTCTCTC---AAATCCAGCAAGAGTCTCTCAAGAGATGGGATACCAAGTT 1535
Qy      1523 TTGAAATCAATCAAGAAAGAGCCAGGCACTAAGTGTGCTGTTGATGATCAAAACT 1582
Db      1536 TTGAGCAGTCCAAAGAAAGAGGCTTGCGGCTGAGGCACTGTTGAAAGAGAGCAAGT 1595
Qy      1583 TTTCACTCTCAAGATTTTGGGCTGATGATTTCAAGCACTGATTTCAAAATGATCAGAA 1641
Db      1596 TTCAACTCAAGATTTTGGATGCTGTATTTCAAAACAATTAATCAACTGTAAACGAA 1654

```

```

ACCESSION AC115775.5 GI:28475880
VERSION HTG1_PHA5E1; HTGS_DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 191072)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-125P1
Unpublished
2 (bases 1 to 191072)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barra, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Bouhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamet, A., Karatas, A., Kells, C., Lacroque, K., Larzates, R.,
Lander, T., Lenoczky, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Mathews, C.,
McCarthy, M., McEwan, P., McKernan, K., Melidrim, J., Menus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, J., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Tsigililo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 191072)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, S., Arachchi, H.M., Barra, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cooke, P., Cotum, B., Dearellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafey, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamet, A., Karatas, A., Kells, C., Lander, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Melidrim, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-FEB-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 24, 2003 this sequence version replaced gi:28460925.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23195
Center clone name: 125_P_1

```

RESULT 15  
 AC115775/c  
 LOCUS  
 DEFINITION Mus musculus clone RP23-125P1, WORKING DRAFT SEQUENCE, 12 unordered  
 pieces.



Mon Sep 13 09:12:36 2004

us-10-080-960-1.rge

Page 20

Job time : 4450.98 secs

---





PT brain disorders.

XX Claim 1, Fig 1, 184pp; English.

XX This invention relates to the DNA and protein sequences of novel isolated  
CC polymucleotides 80090, 52874, 52880, 63497, or 33425 encoding novel  
CC fucosyl transferase, seven transmembrane receptor, or Rho GTPase-  
CC activating protein (RhoGAP) family members. The DNA and protein sequences  
CC of the invention are useful for identifying a compound capable of  
CC treating a disorder by aberrant expression of the nucleic acid or the  
CC activity of the protein, in evaluating the efficacy of treatments and in  
CC diagnosis. The invention also comprises a method for treating a disorder  
CC using a modulator of the proteins activity. This method is useful for  
CC treating breast, ovarian, prostate, colon, or lung cancer, heart,  
CC cardiovascular (including endothelial cell), haematopoietic, blood  
CC vessel and brain disorders. It is also useful in treating pain and  
CC metabolic disorders, liver disorders and platelet disorders. The DNA and  
CC protein sequences are also useful in treating proliferative disorders  
CC (e.g. acute leukaemia), cardiovascular disorders (e.g. cardiac  
CC hypertrophy, myocardial infarction, rheumatic fever), endothelial cell  
CC disorders (e.g. psoriasis, Grave's disease, rheumatoid arthritis),  
CC haematopoietic disorders (e.g. diabetes mellitus, multiple sclerosis,  
CC Crohn's disease, asthma), blood vessel disorders (e.g. Kaposi sarcoma),  
CC degenerative disorders (e.g. Alzheimer disease), and hormonal disorders.  
CC The sequences of the invention are also useful as query sequences to  
CC identify other family members or related sequences. The present sequence  
CC represents the cDNA sequence encoding the human 80090 protein of the  
CC invention  
XX

SO Sequence 1669 BP; 480 A; 375 C; 378 G; 436 T; 0 U; 0 Other;

Query Match 99.8%; Score 1665.8; DB 6; Length 1669;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1667; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CACGGCTCGGCTGTGCTCTAGTGTGACTTTGGCGCTCAGGTATCCAGCTTT 60  
DB 1 CAGCGTCGGCTGTGCTCTAGTGTGACTTTGGCGCTCAGGTATCCAGCTTT 60  
QY 61 TAAAGCCAATATAATTTCTTACTCCTTGAGTGTCTGCTGCTTCACTAGTGT 120  
DB 61 TAAAGCCAATATAATTTCTTACTCCTTGAGTGTCTGCTGCTTCACTAGTGT 120  
QY 121 TTTTCTTTCTTTTGGCTTGTGATCCCTTGAGTCTTAAGAAAGTACGGGCT 180  
DB 121 TTTTCTTTCTTTTGGCTTGTGATCCCTTGAGTCTTAAGAAAGTACGGGCT 180  
QY 181 CCCAGGAGATTACAGACTCCATGCAATGCTTCAATGATGAGTGCCTTTATCAACC 240  
DB 181 CCCAGGAGATTACAGACTCCATGCAATGCTTCAATGATGAGTGCCTTTATCAACC 240  
QY 241 AGAGAGACGAGACACTTAAAGAGTCACTGTTGAGCTGGGAACTTTGAAGAAAGAG 300  
DB 241 AGAGAGACGAGACACTTAAAGAGTCACTGTTGAGCTGGGAACTTTGAAGAAAGAG 300  
QY 301 TTTTAAAGTTCAGTTTGCAAGATGACATCAAAATAGAGAAAGCACTTACCATCTT 360  
DB 301 TTTTAAAGTTCAGTTTGCAAGATGACATCAAAATAGAGAAAGCACTTACCATCTT 360  
QY 361 AATTCATTTCTTAAGAAAGAGATTGACCTTCAACAGAGAAAGAAATGGGAATTGAC 420  
DB 361 AATTCATTTCTTAAGAAAGAGATTGACCTTCAACAGAGAAAGAAATGGGAATTGAC 420  
QY 421 AGCTACCCCATATGCTGTGAGTCCCGCTGAGAGGGGAGACTGGGAGTTTGGGCAA 480  
DB 421 AGCTACCCCATATGCTGTGAGTCCCGCTGAGAGGGGAGACTGGGAGTTTGGGCAA 480  
QY 481 TGTGAGACAGATGCTTGTCTTCAACATCAACCGGACTTACCTTCATCATCATGACC 540  
DB 481 TGTGAGACAGATGCTTGTCTTCAACATCAACCGGACTTACCTTCATCATCATGACC 540  
QY 541 AAAGCTTCCTCTCTATGATGATCTTAACTAGATAGTACTTACCTTGGCTCGGAAA 600  
DB 541 AAAGCTTCCTCTCTATGATGATCTTAACTAGATAGTACTTACCTTGGCTCGGAAA 600

DB 541 AAAGCTTCCTCTCTATGATGATCTTAACTAGATAGTACTTACCTTGGCTCGGAAA 600  
QY 601 GCCCATCATGACTGGGCTGTTTTCATGAAGAGCCCCGAAAGAAATTAAGTCTTT 660  
DB 601 GCCCATCATGACTGGGCTGTTTTCATGAAGAGCCCCGAAAGAAATTAAGTCTTT 660  
QY 661 CATTAACAGAGATTACCTGTTTCACTAGACTGCCACTGTTGACAGGCAATGCCACTTG 720  
DB 661 CATTAACAGAGATTACCTGTTTCACTAGACTGCCACTGTTGACAGGCAATGCCACTTG 720  
QY 721 CCACCTAATCCCAATACCTTGGAGAGCAATTGAAGTCTGAAGTCACTCCGATCTTACT 780  
DB 721 CCACCTAATCCCAATACCTTGGAGAGCAATTGAAGTCTGAAGTCACTCCGATCTTACT 780  
QY 781 CTTTGCAGTCCAAAACAGCTTAGAAAAGACTTCTCGCTGGTGTATGATCAAGTCA 840  
DB 781 CTTTGCAGTCCAAAACAGCTTAGAAAAGACTTCTCGCTGGTGTATGATCAAGTCAAGTCA 840  
QY 841 GACTGTGACCCACATCAGACAGGAGACGATGTTGCGAGCTGATGACTTACATCGAG 900  
DB 841 GACTGTGACCCACATCAGACAGGAGACGATGTTGCGAGCTGATGACTTACATCGAG 900  
QY 901 GTCCATTCCTATGGTGAATGTTTACGAACAAAGACCTCCCTGAGCAGCTGAAATTCGA 960  
DB 901 GTCCATTCCTATGGTGAATGTTTACGAACAAAGACCTCCCTGAGCAGCTGAAATTCGA 960  
QY 961 GCTCTATGATGATGCCATGAGCTTTTATAGATCAATTGACAGATTAAGTTATCTAGCT 1020  
DB 961 GCTCTATGATGATGCCATGAGCTTTTATAGATCAATTGACAGATTAAGTTATCTAGCT 1020  
QY 1021 TTTGAGATGAGATTTTGTGATGATACATACATGAGAAAGTTGCGAGCCACATGAACTG 1080  
DB 1021 TTTGAGATGAGATTTTGTGATGATACATACATGAGAAAGTTGCGAGCCACATGAACTG 1080  
QY 1081 GGGGTAGTCCCTGTATATTAACGATCCCGACAGATCAGACATGCGCTTCCAAATGAA 1140  
DB 1081 GGGGTAGTCCCTGTATATTAACGATCCCGACAGATCAGACATGCGCTTCCAAATGAA 1140  
QY 1141 AGTGTATTTCTTGTATCAAGATTTTCTCACCCGAGGAACTGGCAAGTTACATCAAGCA 1200  
DB 1141 AGTGTATTTCTTGTATCAAGATTTTCTCACCCGAGGAACTGGCAAGTTACATCAAGCA 1200  
QY 1201 CTGATCTGATGACAGATGATGATGAGGCTATGATAGAAAGTGAAGGCTGATGATC 1260  
DB 1201 CTGATCTGATGACAGATGATGATGAGGCTATGATAGAAAGTGAAGGCTGATGATC 1260  
QY 1261 TCTAACGAGCACTTGTGACAGCTCTCAGGAAAAGGAAATGGGAGTGCAGAGCTCAAC 1320  
DB 1261 TCTAACGAGCACTTGTGACAGCTCTCAGGAAAAGGAAATGGGAGTGCAGAGCTCAAC 1320  
QY 1321 CAGGACATTAATCATCGATGCACTTGTGATGATGATGATGATGATGATGATGATGATG 1380  
DB 1321 CAGGACATTAATCATCGATGCACTTGTGATGATGATGATGATGATGATGATGATGATG 1380  
QY 1381 AGGCTTCAGAAAAGGGCTTACACCCCAAAAGATGGGAGCAGAAATACCACTGAGT 1440  
DB 1381 AGGCTTCAGAAAAGGGCTTACACCCCAAAAGATGGGAGCAGAAATACCACTGAGT 1440  
QY 1441 TCCCAAGAGCCCAAGTGTGCTTCTCACCACTCCGACCTCACCTTGTAGCTTTTG 1500  
DB 1441 TCCCAAGAGCCCAAGTGTGCTTCTCACCACTCCGACCTCACCTTGTAGCTTTTG 1500  
QY 1501 CGAGAGATGATGATTTTCAAGCTTGTGAACAATCCAGAAAGAAACCCAGGACCTAAGTGG 1560  
DB 1501 CGAGAGATGATGATTTTCAAGCTTGTGAACAATCCAGAAAGAAACCCAGGACCTAAGTGG 1560  
QY 1561 CTGGTTGATAGAAATCAAAATCTTTCATCTCAAGAGTTTGGGGCTAGTATTCAGAGAC 1620  
DB 1561 CTGGTTGATAGAAATCAAAATCTTTCATCTCAAGAGTTTGGGGCTAGTATTCAGAGAC 1620  
QY 1621 TGATTTCAAAAATGATCAGATGAAACAGAAAAAAGAAAAAAGAAAAA 1669  
DB 1621 TGATTTCAAAAATGATCAGATGAAACAGAAAAAAGAAAAAAGAAAAA 1669

RESULT 2  
AAH64767 standard; cDNA; 1942 BP.  
ID AAH64767  
XX  
XX AAH64767;  
AC  
XX 11-SEP-2001 (first entry)  
XX  
XX Human secreted protein cDNA, SEQ ID NO: 43.  
DE  
XX Human secreted protein; gene therapy; vaccine; treatment; diagnosis;  
KM GENSET; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200142451-A2.  
PN  
XX  
XX 14-JUN-2001.  
PD  
XX 07-DEC-2000; 2000MO-IB001938.  
PF  
XX 08-DEC-1999; 99US-0169629P.  
PR 06-MAR-2000; 2000US-0187470P.  
XX  
XX (GENSET) GENSET.  
PA  
XX  
XX Dumas Milne Edwards J, Bouguenieret L, Jobert S;  
PI WPI; 2001-367870/38.  
XX  
XX P-PSDB; AAG89164.  
DR  
XX  
XX  
XX Full length GENSET human nucleic acids encoding potentially secreted  
PT proteins, useful in gene therapy and vaccination against a variety of  
PT diseases, and for diagnosis of those diseases.  
PS  
XX  
XX Claim 7; Page 606-607; 921pp; English.  
CC The invention relates to full length GENSET human nucleic acids encoding  
CC potentially secreted proteins. The nucleic acids and the polypeptides  
CC they encode may be used in the prevention, treatment and diagnosis of  
CC diseases associated with inappropriate GENSET gene expression. For  
CC example, they be used to treat disorders associated with decreased GENSET  
CC gene expression by rectifying mutations or deletions in a patient's  
CC genome that affect the activity of GENSET or by supplementing the  
CC patient's own production of GENSET polypeptides. Conversely, antisense  
CC nucleic acid molecules may be administered to down regulate GENSET  
CC expression by binding with the cells' own genes and preventing their  
CC expression. The sense and antisense nucleic acids may also be used as DNA  
CC probes in diagnostic assays to detect and quantitate the presence of  
CC similar nucleic acid sequences in samples, and hence to determine which  
CC patients may be in need of restorative therapy. The GENSET polypeptides  
CC may be used as antigens in the production of antibodies and in assays to  
CC identify modulators (agonists and antagonists) of GENSET polypeptide  
CC expression and activity. The present sequence is a GENSET nucleic acid of  
CC the invention  
SQ Sequence 1942 BP; 545 A; 460 C; 462 G; 475 T; 0 U; 0 Other;  
Query Match 90.9%; Score 1517.2; DB 5; Length 1942;  
Best Local Similarity 99.1%; Freq. No. 0;  
Matches 1535; Conservative 0; Mismatches 13; Indels 1; Gaps 1;  
QY 118 TTTTCTTTTCTTTTCTTTTCTTTGCTGATACCGTTGAGATCTAATGAAGCAGGGC 177  
Db 394 TTTTCGCTTGTACACACTCCAGCCTTGATACCGTTAGATCTAATGAAGCAGGGC 453  
QY 178 CTTCCCGCAGGAGTTACAGACTCATGCAATGCTTCATATGATAGTGGCTTTATCTAAC 237  
Db 454 CTTCCCGCAGGAGTTACAGACTCATGCAATGCTTCATATGATAGTGGCTTTATCTAAC 513  
QY 238 ACCAGGACGACGACCATTAAGAGGTGATGCTGGGGAAGTTGAAGAG 297

Db 514 ACCAGGACGACGACCATTAAGAGGTGATGCTGGGGAAGTTGAAGAG 573  
QY 298 GAGTTTAAATTTCCAGTTTCCAGATGACATACAAATAATGAGAGACCTACGCAT 357  
Db 574 GAGTTTAAATTTCCAGTTTCCAGATGACATACAAATAATGAGAGACCTACGCAT 633  
QY 358 CTTAATTCATTTCTTAAAGAAAGATTACCTTCAACAGAAAAAGAAATGGGAATTG 417  
Db 634 CTTAATTCATTTCTTAAAGAAAGATTACCTTCAACAGAAAAAGAAATGGGAATTG 693  
QY 418 GACAGCTTACCCCATTTATGCTGTGTTGCTCCCGGTGACGCGGGAGACTGGAGTTAGGC 477  
Db 694 GACAGCTTACCCCATTTATGCTGTGTTGCTCCCGGTGACGCGGGAGACTGGAGTTAGGC 753  
QY 478 CAATGTGAGAGAGATGCTGTTTCTTCAACATCAACCGGACCTACCTCATCATG 537  
Db 754 CAATGTGAGAGAGATGCTGTTTCTTCAACATCAACCGGACCTACCTCATCATG 813  
QY 538 ACCAAGCATTCCTCTTCTATGCTATGCTACCTTAACTAATAGTTACTCTGCTCGG 597  
Db 814 ACCAAGCATTCCTCTTCTATGCTATGCTACCTTAACTAATAGTTACTCTGCTCGG 873  
QY 598 AAAGCCCATCTGACTGGGCTTTTTCATGAGAGTCCCGGAAAAACAATTATAGCTC 657  
Db 874 AAAGCCCATCTGACTGGGCTTTTTCATGAGAGTCCCGGAAAAACAATTATAGCTC 933  
QY 658 TTTGATTAACCAAGATTAACCTTGTTCATCACTGACGTCACGTTTCAGCAGGACTTCCAC 717  
Db 934 TTTGATTAACCAAGATTAACCTTGTTCATCACTGACGTCACGTTTCAGCAGGACTTCCAC 993  
QY 718 TTGCGACATTAACCTGACATTAATCTGAGAGATGGAAGTCTGTAAGTCACTCCGATACCTA 777  
Db 994 TTGCGACATTAACCTGACATTAATCTGAGAGATGGAAGTCTGTAAGTCACTCCGATACCTA 1053  
QY 778 GTTCTTTGACGTCCAAAAACAAGCTTAGAAAAAGCTTCTCGGTGGTATGACAG 837  
Db 1054 GTTCTTTGACGTCCAAAAACAAGCTTAGAAAAAGCTTCTCGGTGGTATGACAG 1113  
QY 838 TCAAGCTGTGACCCACATCAAGACAGGAGACGTATGTTGCGAGCTGATGACTTACATC 897  
Db 1114 TCAAGCTGTGACCCACATCAAGACAGGAGACGTATGTTGCGAGCTGATGACTTACATC 1173  
QY 898 GAGGTGCGATTCCTTGTGTAATGTTTACGAAACAAAGCCTCCCTGAGAGCTGAAAAAT 957  
Db 1174 GAGGTGCGATTCCTTGTGTAATGTTTACGAAACAAAGCCTCCCTGAGAGCTGAAAAAT 1233  
QY 958 CCAGCTCTATGATGCGAGTGGCTTTTATGAGATCATGCAAGTATTAATCTTA 1017  
Db 1234 CCAGCTCTATGATGCGAGTGGCTTTTATGAGATCATGCAAGTATTAATCTTA 1293  
QY 1018 GCTTTGAGATGAGATTTGTGATGATCAATCACTGAGAAAGTTCTGGAGCCACTGAAA 1077  
Db 1294 GCTTTGAGATGAGATTTGTGATGATCAATCACTGAGAAAGTTCTGGAGCCACTGAAA 1353  
QY 1078 CTGGGGGTAGTCCCTGTATATTACGATCCCGACGATCAAGACTGGCTTCCAACTAAC 1137  
Db 1354 CTGGGGGTAGTCCCTGTATATTACGATCCCGACGATCAAGACTGGCTTCCAACTAAC 1413  
QY 1138 AAAAGTCTATTCTTGTATGAAATTTTCTACCCGAGGAACTGGCAATTCATCAAA 1197  
Db 1414 AAAAGTCTATTCTTGTATGAAATTTTCTACCCGAGGAACTGGCAATTCATCAAA 1473  
QY 1198 CGACTGATTTCTGATGACAGATTGTATGAGGCTTATGTAATGAGAGCTGAAAGGCTGAG 1257  
Db 1474 CGACTGATTTCTGATGACAGATTGTATGAGGCTTATGTAATGAGAGCTGAAAGGCT AG 1532  
QY 1258 ATCTTAACGAGGACTTCTGACAGCTTCTGAGGAAAGGAAATGGGAGTGAAGAGCTC 1317  
Db 1533 ATCTTAACGAGGACTTCTGACAGCTTCTGAGGAAAGGAAATGGGAGTGAAGAGAGCTC 1592  
QY 1318 AACGAGACAAATTAACATGATGATTTGAGTGTATGATGACCAAGAGTGGGCTAAT 1377

Db 1593 AACGAGACAAATTACATCGATCTGATTTGATGTGTGTCACCAAGGTGTGGGCTAAT 1652  
Qy 1378 ATCAGGCTTCAGGAAAAGGGCTTACCCACCAAAAGATGGAGGAGAGATACCCACTG 1437  
Db 1653 ATAGGCTTCAGGAAAAGGGCTTACCCACCAAAAGATGGAGGAGAGATACCCACTG 1712  
Qy 1438 AGTGGCCAGAGCCCAAGATGTTGCTTTTCACCACTCCGAGCTTCCACCTTGAAGCT 1497  
Db 1713 AGTTGCCAGAGCCCAAGATGTTGCTTTTCACCACTCCGAGCTTCCACCTTGAAGCT 1772  
Qy 1498 TTGCGAGAGATGAGATTTCCAGCTTTGACATCCAGAAAGAGCCGAGCACTAAG 1557  
Db 1773 TTGCGAGAGATGAGATTTCCAGCTTTGACATCCAGAAAGAGCCGAGCACTAAG 1832  
Qy 1558 TGGCTGTTGATAGAGATCAAACTTTTCATCTCAAGAGTTTGGGGCTTATGTTCAAG 1617  
Db 1833 TGGCTGTTGATAGAGATCAAACTTTTCATCTCAAGAGTTTGGGGCTTATGTTCAAG 1892  
Qy 1618 GACTGATTTCAAAAATGATCAGATGAAACGAGAAAAAAGAAAAA 1667  
Db 1893 GACTGATTTCAAAAATGATCAGATGAAACGAGAAAAAAGAAAAA 1942

RESULT 3

AD30307  
ID AD30307 standard; cDNA; 2304 BP.

AC AD30307;  
DT 18-DEC-2003 (first entry)  
DE Human novel cDNA sequence, SEQ ID NO:389.

XX Human; diagnostic; drug screening; forensics; gene mapping;  
XX biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
XX neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
XX ulcers; osteoporosis; autoimmune disease; cancer;  
XX molecular weight marker; food supplement; antiparkinsonian; nocitropic;  
XX neuroprotective; anti-nausea; anticoagulant; thrombolytic; vulnery;  
XX antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
XX gene therapy; chromosome 8; gene; ss.

OS Homo sapiens.

PN MO2003029271-A2.

XX 10-APR-2003.

PF 24-SEP-2002; 2002MO-US030474.

XX 24-SEP-2001; 2001US-0324631P.

PA (HYSEQ-) HYSEQ INC.

XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Drmanac R;

XX WPI; 2003-371981/35.  
DR P-PSDB; AD3031278.

XX New polynucleotide and polypeptide useful for diagnosing, preventing or  
PT treating conditions such as neurodegenerative diseases, anemias, platelet  
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
PT cancer.

PS Claim 1; SEQ ID NO 389; 1185bp; English.

XX The invention relates to 971 novel human cDNA sequences (AD3029919-  
CC AD30889) and the polypeptides they encode (AD30890-AD30860). The  
CC invention also relates to nucleic acid sequences over 99% identical with  
CC the novel human cDNAs. The invention additionally encompasses expression  
CC vectors and host cells comprising a nucleic acid of the invention; the

CC recombinant production of a polypeptide of the invention; an antibody  
CC against a polypeptide of the invention; a method of detecting  
CC polynucleotides or polypeptides of the invention; and methods of  
CC identifying a compound which binds to a polypeptide of the invention. The  
CC invention further discloses methods of preventing, treating or  
CC ameliorating a medical condition, kits comprising polynucleotide probes  
CC and/or monoclonal antibodies for carrying out the methods of the  
CC invention; methods for the identification of compounds that modulate the  
CC expression or activity of the polynucleotide and/or polypeptide; and 767  
CC coding sequences corresponding to the cDNA sequences of the invention  
CC (AD303161-AD303267) and the polypeptides encoded by the contigs (AD303268  
CC -AD303394). The nucleic acids and polypeptides of the invention are  
CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
CC identification of mutations responsible for genetic disorders or other  
CC traits, for assessing biodiversity, and in producing many other types of  
CC data and products dependent on DNA and amino acid sequences. They are  
CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
CC disease and other neurodegenerative diseases, anaemia, platelet  
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
CC cancer. The nucleic acids may also be used as hybridisation probes or  
CC primers, and in the recombinant production of a protein. The polypeptides  
CC are also useful in generating antibodies, as molecular weight markers,  
CC and as food supplements. The present sequence represents a specifically  
CC claimed human cDNA sequence of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 2304 BP; 639 A; 521 C; 535 G; 609 T; 0 U; 0 Other;

XX Query Match 83.0%; Score 1385.8; DB 9; Length 2304;  
XX Best Local Similarity 99.5%; Pred. No. 0;  
XX Matches 1390; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 259 AAGAGGTCATGTTGAGCTGGGAGTTTGAAGAGAGTTTAAAGTTCCAGTTG 318  
Db 432 AAGAGTGCATGTTGAGCTGGGAGTTTGAAGAGAGTTTAAAGTTCCAGTTG 491  
Qy 319 CAAGATGACATACAAAATGAGAGAGACCTTACGATTTAATTCATTTCTTAAGAA 378  
Db 492 CAAGATGACATACAAAATGAGAGAGACCTTACGATTTAATTCATTTCTTAAGAA 551  
Qy 379 GAAGATGACCTTCAACAGAAAGAAATGGAAATGGACGCTAACCCATTAGCGC 438  
Db 552 GAAGATGACCTTCAACAGAAAGAAATGGAAATGGACGCTAACCCATTAGCGC 611  
Qy 439 TGGTGGTCCCGCTGACGGGGAGACTGGAGGTTTAAAGCCATGTGGAGCAGATGCTGT 498  
Db 612 TGGTGGTCCCGCTGACGGGGAGACTGGAGGTTTAAAGCCATGTGGAGCAGATGCTGT 671  
Qy 499 TTCTTACATCAACCGGACCTTACCTGATCATATATGACCAAGATTCCTCTTCTAT 558  
Db 672 TTCTTACATCAACCGGACCTTACCTGATCATATATGACCAAGATTCCTCTTCTAT 731  
Qy 559 GGTACTGACTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 618  
Db 732 GGTACTGACTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 791  
Qy 619 GTTTTTCATGAAGAGTCCCGGAAACAAATTAAGTCTTTTCAATAACAGATTTAC 678  
Db 792 GTTTTTCATGAAGAGTCCCGGAAACAAATTAAGTCTTTTCAATAACAGATTTAC 851  
Qy 679 TTGTCATCACTACGCGCAGCTTACAGAGGATTCGCACTTGCACATTAATCCCATAC 738  
Db 852 TTGTCATCACTACGCGCAGCTTACAGAGGATTCGCACTTGCACATTAATCCCATAC 911  
Qy 739 TTGAGAGCATTGAAGTCTTGAAGTCACTCCGATACCTAGTTCTTTTGCAGTCCAAAAC 798  
Db 912 TTGAGAGCATTGAAGTCTTGAAGTCACTCCGATACCTAGTTCTTTTGCAGTCCAAAAC 971  
Qy 799 AAGCTTAAGAAAAGCTTGGCTCCGCTGTGTATATAGTCAGAGCTGAGCCACCATCA 858  
Db 972 AAGCTTAAGAAAAGCTTGGCTCCGCTGTGTATATAGTCAGAGCTGAGCCACCATCA 1031

QY 859 GACAGGAGACGCTATGTTGGCAGCTGATGACTTACATGAGGTCGATTCCTATGCTGAA 918  
 DB 1032 GACAGGAGACGCTATGTTGGCAGCTGATGACTTACATGAGGTCGATTCCTATGCTGAA 1091  
 QY 919 TGTTCACGAAACAAAGACCTCCCTCAGAGAGCTGAAAAATCCAGGCTCATAGGATGCCAT 978  
 DB 1092 TGTTCACGAAACAAAGACCTCCCTCAGAGAGCTGAAAAATCCAGGCTCATAGGATGCCAT 1151  
 QY 979 GGGCTTTATAGATCATTTGCACAGTAAATTTATCTTACGCTTTTGAGAAATGCAATTTGT 1038  
 DB 1152 GGGCTTTATAGATCATTTGCACAGTAAATTTATCTTACGCTTTTGAGAAATGCAATTTGT 1211  
 QY 1039 GATGACTACATCATGAGAGAGTTCTGGAGGCCACATGAAATGGGGGATAGTCCCTGTTAT 1098  
 DB 1212 GATGACTACATCATGAGAGAGTTCTGGAGGCCACATGAAATGGGGGATAGTCCCTGTTAT 1271  
 QY 1099 TACGATCCCCCAGCATCACAGACTGGCTTCCAAATAACAAAGTGTATTTCTGATCA 1158  
 DB 1272 TACGATCCCCCAGCATCACAGACTGGCTTCCAAATAACAAAGTGTATTTCTGATCA 1331  
 QY 1159 GAATTTTCTCACCCAGGAACTGGCAAGTATCATCAGACACTGGATTTCTGATGACAGA 1218  
 DB 1332 GAATTTTCTCACCCAGGAACTGGCAAGTATCATCAGACACTGGATTTCTGATGACAGA 1391  
 QY 1219 TTGTATGAGGCTTATGTAGAAATGGAAGGTGAAGGTGATCTCTTACACGAGCTTTCTG 1278  
 DB 1392 TTGTATGAGGCTTATGTAGAAATGGAAGGTGAAGGTGATCTCTTACACGAGCTTTCTG 1451  
 QY 1279 AAGGCTTTCAGGAAACGGAATGGGGAGTGCAGACGTCACACGACAAATTAATCATGAT 1338  
 DB 1452 AAGGCTTTCAGGAAACGGAATGGGGAGTGCAGACGTCACACGACAAATTAATCATGAT 1511  
 QY 1339 GGATTTGAGTATGATGTGTGACACCAAGGTGAGGCTTAATCATCAGGCTTTCAGGAAAGGGC 1398  
 DB 1512 GGATTTGAGTATGATGTGTGACACCAAGGTGAGGCTTAATCATCAGGCTTTCAGGAAAGGGC 1571  
 QY 1399 TTACCAACCCAAAGATGGGAGGACGAAAGATCCACCTGATGTTGCCAGAGCCACAGATG 1458  
 DB 1572 TTACCAACCCAAAGATGGGAGGACGAAAGATCCACCTGATGTTGCCAGAGCCACAGATG 1631  
 QY 1459 TTTGCTTTCTACCACTCCGACCTCCACCTTTGAGCTTTTGGAGAGAGATGCGATTTCC 1518  
 DB 1632 TTTGCTTTCTACCACTCCGACCTCCACCTTTGAGCTTTTGGAGAGAGATGCGATTTCC 1691  
 QY 1519 AGCTTTGACACATCCAGAAAGAACCCAGGACCTAAGAGTGGCTGTTGATGAGATCAA 1578  
 DB 1692 AGCTTTGACACATCCAGAAAGAACCCAGGACCTAAGAGTGGCTGTTGATGAGATCAA 1751  
 QY 1579 AACTTTTATCTCAAGAGTTTGGGGCTTATGATTTCAAGACTGATTTCAAAAATGATCA 1638  
 DB 1752 AACTTTTATCTCAAGAGTTTGGGGCTTATGATTTCAAGACTGATTTCAAAAATGATCA 1811  
 QY 1639 GAATGAACAGAAAAA 1655  
 DB 1812 GAATGAACAGACTNGA 1828

RESULT 4  
 AEN78606  
 ID AEN78606 standard; cDNA; 597 BP.  
 AC AEN78606;  
 DT 08-JUL-2002 (first entry)  
 DE Human ORF353 cDNA, SEQ ID NO:7105.  
 XX  
 XX Human; ORF, open reading frame; ORF; drug screening; diagnosis;  
 KM disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KM immune modulation; haematopoiesis regulation; tissue growth;  
 KM angiogenesis; activin; inhibin; chemokine; chemokine; haemostatic;  
 KM thrombolytic; tumour inhibition; bodily characteristic; fertility;

KM behaviour; cancer; proliferative disorder; neurological disorder;  
 KM cardiovascular disease; immune system disorder; organ transplantation;  
 KM tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KM hypothyroidism; cholesterol ester storage disease; infection; vulnerability;  
 KM vasculitic; antipsoriatic; antidiabetic; cytostatic; nocotropic;  
 KM neuroprotective; antihypertensive; antidiabetic; antidiabetic; thrombolytic;  
 KM cardiac; hypotensive; antidiabetic; antidiabetic; immunomodulator;  
 KM dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.  
 OS Homo sapiens.  
 XX  
 XX MO200190366-A2.  
 PN  
 XX  
 XX 29-NOV-2001.  
 PD  
 XX  
 XX 24-MAY-2001; 2001MO-US017076.  
 PF  
 XX  
 XX 24-MAY-2000; 2000US-0206690P.  
 PR  
 XX  
 XX (CUBA-) CUBAGEN CORP.  
 PA  
 XX  
 XX Leach MD, Shinkets RA;  
 PI  
 XX  
 XX WPI; 2002-106200/14.  
 DR  
 XX  
 XX P-PSDB; ABP34580.  
 PT  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation.  
 PS  
 XX  
 XX Claim 1; Page 2022; 2508bp; English.

CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences AEN75054-  
 CC AEN79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, tumour inhibition activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antiinfective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases

XX  
 XX Sequence 597 BP; 171 A; 145 C; 124 G; 157 T; 0 U; 0 Other;  
 SO  
 Query Match 34.8%; Score 581.2; DB 6; Length 597;  
 Best Local Similarity 99.3%; Pred. No. 1.3e-150;  
 Matches 594; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

QY 624 TCATGAGAGTCCCGGAAAAAATATTAGCTTTTCATTAACCAAGTATTAAGCTTGT 683
DB 1 TCATGAGAGTCCCGG-AAAAATATTAGCTTTTCATTAACCAAGTATTAAGCTTGT 59
QY 684 CAATGACACTGCGCAGTTCAGCAGCAGTCCCTGCTGCTAACTAACCAATATCTTGA 743
DB 60 CAATGACACTGCGCAGTTCAGCAGCAGTCCCTGCTGCTAACTAACCAATATCTTGA 119
QY 744 GAGCATTTGAAGTCCGAGTCACTCCGACTAGTCTTTCGAGTCCAAAAAAGT 803
DB 120 GAGCATTTGAAGTCCGAGTCACTCCGACTAGTCTTTCGAGTCCAAAAAAGT 179
QY 804 TAGAAAAAGACTGCTCCGCTGGTGTATGTACAGTCACTGTGACCCCATCAGACAG 863
DB 180 TAGAAAAAGACTGCTCCGCTGGTGTATGTACAGTCACTGTGACCCCATCAGACAG 239
QY 864 GAGCAGCTATGTTCCGAGCTCATGACTTACATGAGGTGATTCCTATGATGATGTT 923
DB 240 GAGCAGCTATGTTCCGAGCTCATGACTTACATGAGGTGATTCCTATGATGATGTT 299
QY 924 ACGAAACAAGACCTCCCTCAGCAGCTGAAAAATCCAGCTCTATGATGATGCGATGCTT 983
DB 300 ACGAAACAAGACCTCCCTCAGCAGCTGAAAAATCCAGCTCTATGATGATGCGATGCTT 359
QY 984 TTATAGCATTTGACAGATTAAGTTATCTTACGCTTTTGAAGATGCACTTTGTATGA 1043
DB 360 TTATAGCATTTGACAGATTAAGTTATCTTACGCTTTTGAAGATGCACTTTGTATGA 419
QY 1044 CTACATCAGTGAAGATTTCTGAGGCGCACTGAAAATGCGGGGAGTCCCTGATATTCAG 1103
DB 420 CTACATCAGTGAAGATTTCTGAGGCGCACTGAAAATGCGGGGAGTCCCTGATATTCAG 479
QY 1104 ATCCCCAGCATCAGACTGCTTCCAAATGACAAAAGTCTATTTCTTATGATGAT 1163
DB 480 ATCCCCAGCATCAGACTGCTTCCAAATGACAAAAGTCTATTTCTTATGATGAT 539
QY 1164 TTCTCAGCCAGGAGAACTGGCAGTATCACTGAGAGTGTGATGTCAGACATG 1221
DB 540 TTCTCAGCCAGGAGAACTGGCAGTATCACTGAGAGTGTGATGTCAGACATG 597

```

RESULT 5  
ADCC32194/c  
ID ADCC32194 standard; cDNA; 447 BP.

AC ADCC32194;  
AT  
AT  
18-DEC-2003 (First entry)  
DE Human novel cDNA contig sequence, SEQ ID NO:2276.  
XX  
XX  
XX Human; diagnostic; drug screening; forensics; gene mapping;  
XX blood; diversity assessment; Parkinson's disease; Alzheimer's disease;  
XX neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
XX ulcers; osteoporosis; autoimmune disease; cancer;  
XX molecular weight marker; food supplement; antiParkinsonian; nocitopic;  
XX neuroprotective; antianemic; anticoagulant; thrombolytic; vulnerytic;  
XX anticancer; osteoparitic; immunosuppressive; antiinflammatory; cytostatic;  
XX gene therapy; chromosome 8; ss.  
OS Homo sapiens.  
PN MO2003029271-A2.  
PD 10-APR-2003.  
XX  
XX 24-SEP-2002; 2002MO-US030474.  
PF  
XX 24-SEP-2001; 2001US-0324631P.  
PR  
XX (HYSE-) HYSEQ INC.  
PA

PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Drmanac RT;  
XX  
DR WPI: 2003-371981/35.  
DR P-PSDB; ADCC32961.  
XX  
XX New polynucleotide and polypeptide useful for diagnosing, preventing or  
PT treating conditions such as neurodegenerative diseases, anemias, platelet  
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
PT cancer.  
XX  
XX Example 2; SEQ ID NO 2276; 1185bp; English.  
PS  
XX The invention relates to 971 novel human cDNA sequences (ADCC32919-  
CC ADCC30889) and the polypeptides they encode (ADCC30890-ADCC31860). The  
CC invention also relates to nucleic acid sequences over 99% identical with  
CC the novel human cDNAs. The invention additionally encompasses expression  
CC vectors and host cells comprising a nucleic acid of the invention; the  
CC recombinant production of a polypeptide of the invention; an antibody  
CC against a polypeptide of the invention; a method of detecting  
CC polynucleotides or polypeptides of the invention; and methods of  
CC identifying a compound which binds to a polypeptide of the invention. The  
CC invention further discloses methods of preventing, treating or  
CC ameliorating a medical condition; kits comprising polynucleotide probes  
CC and/or monoclonal antibodies for carrying out the methods of the  
CC invention; methods for the identification of compounds that modulate the  
CC expression or activity of the polynucleotide and/or polypeptide; and 767  
CC contig sequences corresponding to the cDNA sequences of the invention  
CC (ADCC31861-ADCC32627) and the polypeptides encoded by the contigs (ADCC32628  
CC -ADCC33394). The nucleic acids and polypeptides of the invention are  
CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
CC identification of mutations responsible for genetic disorders or other  
CC traits, for assessing biodiversity, and in producing many other types of  
CC data and products dependent on DNA and amino acid sequences. They are  
CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
CC disease and other neurodegenerative diseases, anaemia, platelet  
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
CC cancer. The nucleic acids may also be used as hybridisation probes or  
CC primers, and in the recombinant production of a protein. The polypeptides  
CC are also useful in generating antibodies, as molecular weight markers,  
CC and as food supplements. The present sequence represents a human contig  
CC sequence used in an example of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
XX  
SQ Sequence 447 BP; 113 A; 103 C; 101 G; 130 T; 0 U; 0 Other;

Query Match 24.9%; Score 416.4; DB 9; Length 447;  
Best Local Similarity 99.8%; Pred. No. 5.9e-105;  
Matches 417; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 795 AAAAAGCTTNGAAAAAGCTTCTCCGCTGGTGTATGACGTACAGTGGACCCAC 853  
DB 447 AAAAAGCTTNGAAAAAGCTTCTCCGCTGGTGTATGACGTACAGTGGACCCAC 388  
QY 854 CATCAGACAGGAGACGATATGTTGGCAGCTGATGACTTTCATCGAGAGTTCATG 913  
DB 387 CATCAGACAGGAGACGATATGTTGGCAGCTGATGACTTTCATCGAGAGTTCATG 328  
QY 914 GTGAATGTTTACGAAACAAAGACTCCCTCAGCAGCTGAAAAATCCAGCCTCATGAG 973  
DB 327 GTGAATGTTTACGAAACAAAGACTCCCTCAGCAGCTGAAAAATCCAGCCTCATGAG 268  
QY 974 CCGATGGCTTTTATAGATCATTCGACAGTATAAGTTATCTAGCTTTTGAAGATGAG 1033  
DB 267 CCGATGGCTTTTATAGATCATTCGACAGTATAAGTTATCTAGCTTTTGAAGATGAG 208  
QY 1034 TTTGTGATGACTACATCAGTGAAGTTCTGGAGGCCATGAAATCGGGGGTGTGCCCG 1093  
DB 207 TTTGTGATGACTACATCAGTGAAGTTCTGGAGGCCATGAAATCGGGGGTGTGCCCG 148

```
QY      1094 TATATTACGATCCCCGATCAACAGCTTGCTTCCAGTAACAAAGTCTATTCTTG 1153
      |||
Db      147 TATATTACGATCCCCGATCAACAGCTTGCTTCCAGTAACAAAGTCTATTCTTG 88
      |||
QY      1154 TATCGAATTTTCTACACCGAGGAACTGCACTTACATCAGACGACTGATTCTGA 1211
      |||
Db      87 TATCGAATTTTCTACACCGAGGAACTGCACTTACATCAGACGACTGATTCTGA 30
      |||

RESULT 6
AAL34845
ID      AAL34845 standard; cDNA, 378 BP.
XX
AC      AAL34845;
XX
DT      08-JAN-2002 (first entry)
XX
DE      Human musculoskeletal system related polynucleotide SEQ ID NO 187.
XX
XX      Cytostatic; immunosuppressive; nocotropic; neuroprotective; antiviral;
KW      antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW      vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW      cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW      neurological disease; infection; human; secreted protein;
KW      musculoskeletal system; ss.
XX
XX      Homo sapiens.
XX      OS
XX      PN      WO200155367-A1.
XX      PD      02-AUG-2001.
XX      PF      17-JAN-2001; 2001WO-US001338.
XX
PR      31-JAN-2000; 2000US-0179065P.
PR      04-FEB-2000; 2000US-0180628P.
PR      24-FEB-2000; 2000US-0184664P.
PR      02-MAR-2000; 2000US-0186350P.
PR      16-MAR-2000; 2000US-0189874P.
PR      17-MAR-2000; 2000US-0190076P.
PR      18-APR-2000; 2000US-0198123P.
PR      19-MAY-2000; 2000US-0205515P.
PR      07-JUN-2000; 2000US-0209467P.
PR      28-JUN-2000; 2000US-0214886P.
PR      30-JUN-2000; 2000US-0215135P.
PR      07-JUL-2000; 2000US-0216647P.
PR      07-JUL-2000; 2000US-0216880P.
PR      11-JUL-2000; 2000US-0217487P.
PR      14-JUL-2000; 2000US-0217496P.
PR      14-JUL-2000; 2000US-0218290P.
PR      26-JUL-2000; 2000US-0220963P.
PR      14-AUG-2000; 2000US-0220964P.
PR      14-AUG-2000; 2000US-0224518P.
PR      14-AUG-2000; 2000US-0224519P.
PR      14-AUG-2000; 2000US-0225213P.
PR      14-AUG-2000; 2000US-0225214P.
PR      14-AUG-2000; 2000US-0225266P.
PR      14-AUG-2000; 2000US-0225267P.
PR      14-AUG-2000; 2000US-0225268P.
PR      14-AUG-2000; 2000US-0225270P.
PR      14-AUG-2000; 2000US-0225447P.
PR      14-AUG-2000; 2000US-0225757P.
PR      14-AUG-2000; 2000US-0225758P.
PR      14-AUG-2000; 2000US-0225759P.
PR      18-AUG-2000; 2000US-0226279P.
PR      22-AUG-2000; 2000US-0226681P.
PR      22-AUG-2000; 2000US-0226686P.
PR      22-AUG-2000; 2000US-0227182P.
PR      23-AUG-2000; 2000US-0227009P.
PR      30-AUG-2000; 2000US-0228924P.
PR      01-SEP-2000; 2000US-0229287P.
PR      01-SEP-2000; 2000US-0229343P.
PR      01-SEP-2000; 2000US-0229344P.
PR      01-SEP-2000; 2000US-0229345P.
PR      03-SEP-2000; 2000US-0229509P.
PR      05-SEP-2000; 2000US-0229513P.
PR      06-SEP-2000; 2000US-0230437P.
PR      06-SEP-2000; 2000US-0230438P.
PR      08-SEP-2000; 2000US-0231242P.
PR      08-SEP-2000; 2000US-0231243P.
PR      08-SEP-2000; 2000US-0231244P.
PR      08-SEP-2000; 2000US-0231413P.
PR      08-SEP-2000; 2000US-0231414P.
PR      08-SEP-2000; 2000US-0232080P.
PR      08-SEP-2000; 2000US-0232081P.
PR      12-SEP-2000; 2000US-0231968P.
PR      14-SEP-2000; 2000US-0232397P.
PR      14-SEP-2000; 2000US-0232398P.
PR      14-SEP-2000; 2000US-0232399P.
PR      14-SEP-2000; 2000US-0232400P.
PR      14-SEP-2000; 2000US-0232401P.
PR      14-SEP-2000; 2000US-0233063P.
PR      14-SEP-2000; 2000US-0233064P.
PR      14-SEP-2000; 2000US-0233065P.
PR      21-SEP-2000; 2000US-0234223P.
PR      21-SEP-2000; 2000US-0234274P.
PR      25-SEP-2000; 2000US-0234997P.
PR      25-SEP-2000; 2000US-0234998P.
PR      26-SEP-2000; 2000US-0235484P.
PR      27-SEP-2000; 2000US-0235834P.
PR      27-SEP-2000; 2000US-0235836P.
PR      29-SEP-2000; 2000US-0236327P.
PR      29-SEP-2000; 2000US-0236367P.
PR      29-SEP-2000; 2000US-0236368P.
PR      29-SEP-2000; 2000US-0236369P.
PR      29-SEP-2000; 2000US-0236370P.
PR      02-OCT-2000; 2000US-0236802P.
PR      02-OCT-2000; 2000US-0237037P.
PR      02-OCT-2000; 2000US-0237038P.
PR      02-OCT-2000; 2000US-0237039P.
PR      02-OCT-2000; 2000US-0237040P.
PR      13-OCT-2000; 2000US-0239355P.
PR      13-OCT-2000; 2000US-0239357P.
PR      20-OCT-2000; 2000US-0240960P.
PR      20-OCT-2000; 2000US-0241221P.
PR      20-OCT-2000; 2000US-0241785P.
PR      20-OCT-2000; 2000US-0241786P.
PR      20-OCT-2000; 2000US-0241787P.
PR      20-OCT-2000; 2000US-0241808P.
PR      20-OCT-2000; 2000US-0241809P.
PR      20-OCT-2000; 2000US-0241826P.
PR      01-NOV-2000; 2000US-0244617P.
PR      08-NOV-2000; 2000US-0246474P.
PR      08-NOV-2000; 2000US-0246475P.
PR      08-NOV-2000; 2000US-0246476P.
PR      08-NOV-2000; 2000US-0246477P.
PR      08-NOV-2000; 2000US-0246523P.
PR      08-NOV-2000; 2000US-0246524P.
PR      08-NOV-2000; 2000US-0246525P.
PR      08-NOV-2000; 2000US-0246526P.
PR      08-NOV-2000; 2000US-0246527P.
PR      08-NOV-2000; 2000US-0246528P.
PR      08-NOV-2000; 2000US-0246532P.
PR      08-NOV-2000; 2000US-0246609P.
PR      08-NOV-2000; 2000US-0246610P.
PR      08-NOV-2000; 2000US-0246611P.
PR      08-NOV-2000; 2000US-0246613P.
PR      17-NOV-2000; 2000US-0249207P.
PR      17-NOV-2000; 2000US-0249208P.
PR      17-NOV-2000; 2000US-0249209P.
PR      17-NOV-2000; 2000US-0249210P.
PR      17-NOV-2000; 2000US-0249211P.
PR      17-NOV-2000; 2000US-0249212P.
PR      17-NOV-2000; 2000US-0249213P.
PR      17-NOV-2000; 2000US-0249214P.
```

CC The invention relates to novel genes (AAL34669-AAJ37666) and proteins  
XX (ABB03087-ABB04109) associated with the musculoskeletal system useful for  
CC preventing, treating or ameliorating medical conditions e.g. by protein  
CC gene therapy. The genes are isolated from a range of human tissues  
CC disclosed in the specification. The nucleic acids, proteins, antibodies  
CC and (antagonists are useful in the diagnosis, treatment and prevention  
CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the  
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.  
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
CC bacterial, fungal and parasitic infections. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from Wipo at  
CC ftp.wipo.int/pub/published\_pat\_sequences

PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 05-DEC-2000; 2000US-0251579P.  
PR 08-DEC-2000; 2000US-0251565P.  
PR 08-DEC-2000; 2000US-0251566P.  
PR 08-DEC-2000; 2000US-0251567P.  
PR 08-DEC-2000; 2000US-0251569P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 11-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0256796P.

(HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Barash SC, Ruben SM;  
WI: 2001-451937/48.  
DR P-PDSB; ABB03263.  
XX  
PT Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders related to the musculoskeletal system including musculoskeletal  
PT cancers and also for testing and detection e.g. diagnosis.  
XX  
PS Claim 1; SEQ ID NO 187; 781bp + Sequence Listing; English.  
XX

|    | Query Match | Similarity   | Score                                    | 33.8 | DB          | 4 | Length | 378  |
|----|-------------|--|--|------|-------------|---|--------|------|
|    | Best Local  | Similarity   | 96.5%                                    | Pred | No. 2.7e-79 |   |        |      |
|    | Matches     | 363  | Conservative                             | 0    | Mismatches  | 8 | Indels | 5    |
|    |             |  |  |      |             |   |        | Gaps |
|    |             |  |  |      |             |   |        | 3    |
| QY | 694         | GCCACGTTAGCAGGCAATTC   | CCACTTGCACCTAACTAACCCAAATCTTGGAGACATTGAA | 753  |             |   |        |      |
| Db | 1           | GCCACGTTAGCAGGCAATTC   | CCACTTGCACCTAACTAACCCAAATCTTGGAGACATTGAA | 60   |             |   |        |      |
| QY | 754         | GTCCTGAATCACTCCGATACCTAGTCTTTCCTTGGACAGTCCAAAAACAAGCTATAGAAAAAG  | 813                                      |      |             |   |        |      |
| Db | 61          | GTCCTGAATCACTCCGATACCTAGTCTTTCCTTGGACAGTCCAAAAACAAGCTATAGAAAAAG  | 120                                      |      |             |   |        |      |
| QY | 814         | CTTGCTCCGCTGGTGTATGTATACATGTGACATCTGTGACCCACCATTCAGCAAGGACACGTAT | 873                                      |      |             |   |        |      |

| Db       | 12   | CTTGGCTCCGCTGGTGTATGTACAGTCAGACGTGTACCCACATCCAGACAGGACAGCTAT              | 180  |
|----------|------|---|------|
| Qy       | 874  | GTTCGGACACTATACACTTAACATGAGGTGCATTCCTCATGTGTAATGTTTACGAACAA               | 933  |
| Db       | 181  | GTTCGGACACTATACACTTAACATGAGGTGCATTCCTCATGTGTAATGTTTACGAACAA               | 240  |
| Qy       | 934  | GACCTCCCTCCAGACCTGAAAAATCCAGCCCTCATGTGATGCCATGGCTTTATATGATC               | 993  |
| Db       | 241  | GACCTCCCTCCAGACCTGAAAAATCCAGCCCTCATGTGATGCCATGGCTTTATATGATC               | 300  |
| Qy       | 994  | ATTGCACAGATTAAGTTTATCTCTAGCTTTTGAGATGCACTTTGTGATCATCATCT                  | 1053 |
| Db       | 301  | A-TGCACAGATTAAGTTTATCTCTAGCTTTT--GAGATGCACTTTGTGATGA-TACATCACT            | 355  |
| Qy       | 1054 | GAGAGTTCTGGAGGC 1069  |      |
| Db       | 356  | GAGAGTTCTGGAGGC 371   |      |
| RESULT 7 |      |   |      |
| ABX57833 | ID   | ABX57833 standard; cDNA; 378 Bp.  |      |
| XX       | AC   | ABX57833;   |      |
| XX       | DT   | 26-FEB-2003 (first entry)   |      |
| DE       | XX   | cDNA encoding novel human musculoskeletal system antigen #177.            |      |
| KM       | KM   | Gene, ss; musculoskeletal system antigen; cancer; metastasis;             |      |
| KM       | KM   | re-vascularisation; thrombosis; arteriosclerosis; mineral content;        |      |
| KM       | KM   | cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;       |      |
| KM       | KM   | post-operative tissue repair; limb regeneration; neuronal growth;         |      |
| KM       | KM   | neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;     |      |
| KM       | KM   | AIDS-related complex; chondrocyte growth; bone regeneration;              |      |
| KM       | KM   | periodontal regeneration; tissue transport; bone graft; skin aging;       |      |
| KM       | KM   | keratinocyte growth; hair loss; melanocyte growth; cell proliferation;    |      |
| KM       | KM   | cell growth; organ transplant; cell differentiation; body height; weight; |      |
| KM       | KM   | hair colour; eye colour; skin; percentage of adipose tissue;              |      |
| KM       | KM   | pigmentation; cosmetic surgery; metabolism; biorhythm; circadian rhythm;  |      |
| KM       | KM   | depression; tendency for violence; pain; reproductive capability;         |      |
| KM       | KM   | hormone level; endocrine level; appetite; libido; memory; stress;         |      |
| KM       | KM   | storage capability; fat content; lipid content; protein content;          |      |
| KM       | KM   | carbohydrate content; vitamin content; cofactor content;                  |      |
| KM       | KM   | nutritional component.  |      |
| XX       | OS   | Homo sapiens.   |      |
| XX       | PN   | US2002147140-A1.  |      |
| PN       | XX   | 10-OCT-2002.  |      |
| PD       | XX   |   |      |
| XX       | XX   |   |      |
| PF       | XX   | 17-JAN-2001; 2001US-00764877.   |      |
| PR       | XX   |   |      |
| PR       | PR   | 31-JAN-2000; 2000US-0179065P.   |      |
| PR       | PR   | 04-FEB-2000; 2000US-0180628P.   |      |
| PR       | PR   | 28-JUN-2000; 2000US-0214885P.   |      |
| PR       | PR   | 07-JUL-2000; 2000US-0218647P.   |      |
| PR       | PR   | 07-JUL-2000; 2000US-0216800P.   |      |
| PR       | PR   | 11-JUL-2000; 2000US-0217487P.   |      |
| PR       | PR   | 11-JUL-2000; 2000US-0214969P.   |      |
| PR       | PR   | 14-JUL-2000; 2000US-0218290P.   |      |
| PR       | PR   | 26-JUL-2000; 2000US-0220965P.   |      |
| PR       | PR   | 26-JUL-2000; 2000US-0220966P.   |      |
| PR       | PR   | 14-AUG-2000; 2000US-0224518P.   |      |
| PR       | PR   | 14-AUG-2000; 2000US-0224519P.   |      |
| PR       | PR   | 14-AUG-2000; 2000US-0225267P.   |      |
| PR       | PR   | 14-AUG-2000; 2000US-0225268P.   |      |
| PR       | PR   | 14-AUG-2000; 2000US-0225270P.   |      |
| PR       | PR   | 14-AUG-2000; 2000US-0225447P.   |      |
| PR       | PR   | 14-AUG-2000; 2000US-0225757P.   |      |
| PR       | PR   | 14-AUG-2000; 2000US-0225786P.   |      |
| PR       | PR   | 22-AUG-2000; 2000US-0226868P.   |      |



PR 30-AUG-2000; 2000US-0228994P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239535P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0244617P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 17-NOV-2000; 2000US-0249298P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 XX  
 XX  
 PA (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 XX  
 XX WPI: 2003-128199/12.  
 DR P-PSDB; ABU12557.  
 XX  
 PT Isolated nucleic acid molecules encoding musculoskeletal system  
 associated polypeptides, useful for detecting disorders, e.g. cancer.  
 XX  
 PS Claim 1; SEQ ID NO 187; 321bp; English.

CC stress; increases or decreases storage capabilities, fat content, lipid,  
 CC protein, carbohydrate, vitamins, minerals, cofactors or other nutritional  
 CC components. This sequence encodes a novel human musculoskeletal system  
 CC antigen. Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from the US patent office at  
 CC ftp.seqdata.uspto.gov/sequence.html?docID=20020147140  
 XX  
 SQ Sequence 378 BP, 107 A; 90 C; 84 G; 96 T; 0 U; 1 Other;  
 Query Match 19.4%; Score 323.8; DB 7; Length 378;  
 Best Local Similarity 96.5%; Pred. No. 2.7e-79;  
 Matches 363; Conservative 0; Mismatches 8; Indels 5; Gaps 3;  
 QY 694 GCCACGTTGACGAGGATTCGCCACTGCGACATCACTACCACTACTTGAGAGCATTGAA 753  
 DB 1 GCCACGTTGACGAGGATTCGCCACTGCGACATCACTACCACTACTTGAGAGCATTGAA 60  
 QY 754 GTTCGTGAGTCACTCCGATACCTAGTTCCTTTGACGTCCAAAACAGCTTGAAAAAGA 813  
 DB 61 GTCTGAAGTCACTCCGATACCTAGTTCCTTTGACGTCCAAAACAGCTTGAAAAAGA 120  
 QY 814 CTTCCTCCGCTGTGTATGTATGACAGTCACTGACCCACCATCAGACAGGACGCTAT 873  
 DB 121 CTTCCTCCGCTGTGTATGTATGACAGTCACTGACCCACCATCAGACAGGACGCTAT 180  
 QY 874 GTTCGACGCTGATGACTTACATCATGAGTGCATTCCTATGCTGATATGTTTACGAAACAA 933  
 DB 181 GTTCGACGCTGATGACTTACATCATGAGTGCATTCCTATGCTGATATGTTTACGAAACAA 240  
 QY 934 GACCTCCCTGACGAGCTGAAAAATCCAGCCTCTATGATGACGATGGCTTTTATAGATC 993  
 DB 241 GACCTCCCTGACGAGCTGAAAAATCCAGCCTCTATGATGACGATGGCTTTTATAGATC 300  
 QY 994 ATTGCACAGTATTAAGTTATCCATGCTTTGAGATGACGATTTGTGATGATCATCACT 1053  
 DB 301 ATTGCACAGTATTAAGTTATCCATGCTTTGAGATGACGATTTGTGATGATCATCACT 355  
 QY 1054 GAGAAGTTCGTGAGGC 1069  
 DB 356 GAGAAGTTCGTGAGGC 371  
 RESULT 8  
 ADCC30780 standard; cDNA; 1061 BP.  
 ID ADCC30780  
 AC  
 XX  
 AC ADCC30780;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE Human novel cDNA sequence, SEQ ID NO:862.  
 XX  
 KW Human; diagnostic; drug screening; forensics; gene mapping;  
 KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
 KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
 KW ulcers; osteoporosis; autoimmune disease; cancer;  
 KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
 KW neuroprotective; antidiabetic; anticoagulant; thrombolytic; vulnerary;  
 KW antitumor; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
 KW gene therapy; chromosome 8; gene; ss.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO2003029271-A2.  
 XX  
 PD 10-APR-2003.  
 XX  
 PF 24-SEP-2002; 2002WO-US030474.  
 XX  
 PR 24-SEP-2001; 2001US-0324631P.  
 XX  
 PA (HYSE-) HYSEQ INC.

XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QH, Wang J, Wehrman T;  
 PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
 PI Haley-Vicente D, Drmanac RT;  
 XX WPI: 2003-371981/35.  
 DR P-PSDB; ADC31751.  
 XX  
 PT New polynucleotide and polypeptide useful for diagnosing, preventing or  
 PT treating conditions such as neurodegenerative diseases, anemias, platelet  
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 PT cancer.  
 XX  
 PS Claim 1, SEQ ID NO 862; 1185bp; English.  
 XX  
 CC The invention relates to 971 novel human cDNA sequences (ADC29919-  
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
 CC invention also relates to nucleic acid sequences over 99% identical with  
 CC the novel human cDNAs. The invention additionally encompasses expression  
 CC vectors and host cells comprising a nucleic acid of the invention; the  
 CC recombinant production of a polypeptide of the invention; an antibody  
 CC against a polypeptide of the invention; a method of detecting  
 CC polynucleotides or polypeptides of the invention; and methods of  
 CC identifying a compound which binds to a polypeptide of the invention. The  
 CC invention further discloses methods of preventing, treating or  
 CC ameliorating a medical condition; kits comprising polynucleotide probes  
 CC and/or monoclonal antibodies for carrying out the methods of the  
 CC invention; methods for the identification of compounds that modulate the  
 CC expression or activity of the polynucleotide and/or polypeptide; and 767  
 CC coding sequences corresponding to the cDNA sequences of the invention  
 CC (ADC31861-ADC32627) and the polypeptides encoded by the cDNAs (ADC32628  
 CC -ADC33394). The nucleic acids and polypeptides of the invention are  
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
 CC identification of mutations responsible for genetic disorders or other  
 CC traits, for assessing biodiversity, and in producing many other types of  
 CC data and products dependent on DNA and amino acid sequences. They are  
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, anaemia, platelet  
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 CC cancer. The nucleic acids may also be used as hybridisation probes or  
 CC primers, and in the recombinant production of a protein. The polypeptides  
 CC are also useful in generating antibodies, as molecular weight markers,  
 CC and as food supplements. The present sequence represents a specifically  
 CC claimed human cDNA sequence of the invention. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pcf\_sequences.  
 XX  
 XX Sequence 1061 BP; 245 A; 255 C; 299 G; 261 T; 0 U; 1 Other;  
 SQ  
 Query Match 18.0%; Score 300; DB 9; Length 1061;  
 Best Local Similarity 100.0%; Pred. No. 1,7e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 9  
 ABS67785  
 ID ABS67785 standard; DNA; 1550 BP.  
 XX  
 AC ABS67785;  
 XX  
 XX  
 DT 29-NOV-2002 (first entry)  
 XX  
 XX  
 DE Human receptors and membrane associated protein REMAP-7 gene.  
 XX  
 XX Human; receptor and membrane-associated protein; REMAP; atherosclerosis;  
 KW cardiovascular disorder; hypertension; congestive heart failure; oedema;  
 KW aneurysm; angina pectoris; ischaemic heart disease; lung disease; nausea;  
 KW rheumatic heart disease; chronic obstructive pulmonary disease; anaemia;  
 KW emphysema; bronchitis; gastrointestinal disorder; peptic ulcer; asthma;  
 KW Crohn's disease; lipid metabolism disorder; Fabry's disease; dementia;  
 KW pancreatitis; diabetes mellitus; hyperlipidaemia; autoimmune disorder;  
 KW inflammatory disorder; neurological disorder; Creutzfeldt-Jakob disease;  
 KW cell proliferative disorder; protein replacement therapy; adenocarcinoma;  
 KW developmental disorder; metabolic disorder; Alzheimer's disease; stroke;  
 KW Parkinson's disease; multiple sclerosis; Cushing's syndrome; anxiety;  
 KW schizophrenia; Addison's disease; endocrine disorder; gene therapy; gout;  
 KW amnesia; cancer; leukaemia; lymphoma; myeloma; sarcoma; gene; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 PN WO200263006-A2.  
 XX  
 PD 15-AUG-2002.  
 XX  
 PF 05-FEB-2002; 2002WO-US003868.  
 XX  
 PR 06-FEB-2001; 2001US-0267201P.  
 PR 16-FEB-2001; 2001US-0269580P.  
 PR 09-APR-2001; 2001US-0282679P.  
 PR 02-MAY-2001; 2001US-0288295P.  
 PR 14-JAN-2002; 2002US-0348687P.  
 XX  
 XX (INCY)- INCYTE GENOMICS INC.  
 PA  
 PI Azinza Y, Yue H, Ding L, Nguyen DB, Burtford N;  
 PI Thangavelu K, Elliott VS, Ramkumar J, Yao MG, Lai PG, Tang TY;  
 PI Swarnakar A, Warren BA, Wallia NK, Policky JL, Xu Y, Homnell CP;  
 PI Au-Young J, Baughn MR, Duggan BM, Lu DM, Gietzen KU, Hillmann UL;  
 PI Raumann BE, Lu Y, Sapperstein SK, Tran UK, Richardson TW;  
 PI Emerling BM, Hafalia AJA, Burrill JD, Marcus GA, Zingler KA;  
 PI Kabie AE, Goward AE;  
 XX  
 XX WPI: 2002-627559/67.  
 DR P-PSDB; ABG92064.  
 DR  
 PT New human receptor and membrane-associated proteins and polynucleotides,  
 PT useful for diagnosing, treating or preventing cardiovascular disorders  
 PT (e.g. aneurysms), neurological disorders (e.g. Parkinson's disease) or  
 PT cancers.  
 XX  
 PS Claim 5; Page 232-233; 262bp; English.  
 XX  
 CC The present invention relates to a new receptor and membrane-associated  
 CC protein (REMAP). The polypeptide, polynucleotide and agonist are useful  
 CC for treating a condition associated with decreased expression of  
 CC functional REMAP. The antagonist is useful for treating a disease  
 CC associated with overexpression of functional REMAP. The anti-REMAP  
 CC antibody is useful for diagnosing a condition or disease associated with  
 CC the expression of REMAP. These polypeptides, polynucleotides, agonists  
 CC and antagonists are particularly useful for diagnosing, treating or  
 CC preventing cardiovascular (e.g. atherosclerosis, hypertension, aneurysms,  
 CC congestive heart failure, angina pectoris, or ischaemic or rheumatic  
 CC heart disease), lung (e.g. oedema, chronic obstructive pulmonary disease,  
 CC emphysema or bronchitis) gastrointestinal (e.g. nausea, peptic ulcer or  
 CC Crohn's disease), lipid metabolism (e.g. Fabry's disease, diabetes

CC mellitus or hyperlipidaemia), autoimmune/inflammatory (e.g. anaemia,  
CC asthma, gout, pancreatitis or Crohn's disease), neurological (e.g.  
CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,  
CC multiple sclerosis, Parkinson's disease, anxiety, schizophrenia or  
CC amnesia), metabolic (e.g. Addison's disease), developmental (e.g.  
CC Cushing's syndrome), endocrine or cell proliferative disorders (e.g.  
CC cancers including adenocarcinoma, leukaemia, lymphoma, myeloma or  
CC sarcoma). The present nucleic acid sequence encodes a human REMAP protein  
CC of the invention  
XX

XX Sequence 1550 BP; 413 A; 344 C; 348 G; 445 T; 0 U; 0 Other;

XX Query Match 17.9%; Score 298; DB 6; Length 1550;  
XX Best Local Similarity 100.0%; Pred. No. 7.4e-72;

XX Matches 298; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 AGGTGATGTTGAGCTGGGAGATTGAAAGAGAGGTTTAAAGTTCCAGTTTGCAG 322  
DB 185 AGGTGATGTTGAGCTGGGAGATTGAAAGAGAGGTTTAAAGTTCCAGTTTGCAG 244  
QY 333 ATGGACATACAAAATGAGAGACACCTACGATCTTAATTCATTCTTAAGAAAGAG 382  
DB 245 ATGGACATACAAAATGAGAGACACCTACGATCTTAATTCATTCTTAAGAAAGAG 304  
QY 383 GATTGACCTTCAACAGAAAGAAATGGAGATTGACAGCTACCCCATTTAGCTTGGT 442  
DB 305 GATTGACCTTCAACAGAAAGAAATGGAGATTGACAGCTACCCCATTTAGCTTGGT 364  
QY 443 GGTCCCGCCCTGACGGGGAGAGCTGGGAGGTTAGGCGCAATGGAGAGAGCTTTGCT 502  
DB 365 GGTCCCGCCCTGACGGGGAGAGCTGGGAGGTTAGGCGCAATGGAGAGAGCTTTGCT 424  
QY 503 TCACCATCAACCGGACCTACCTCATCATCATGACCAAGCAATTCCTCTTATAG 560  
DB 425 TCACCATCAACCGGACCTACCTCATCATCATGACCAAGCAATTCCTCTTATAG 482

RESULT 10  
ABO5334  
ID ABO5334 standard; cDNA; 1479 BP.

AC ABO5334;

DT 26-FEB-2002 (first entry)

DE Human fucosyltransferase family member 32132 coding sequence.

XX Human: fucosyltransferase family member 32132; cytosolic; antitumour;  
XX antiatherosclerotic; hypotensive; antiarrhythmic; vasotrophic;  
XX antidiabetic; antiarrhythmic; immunosuppressive; antihypoid; antitumor;  
XX antiinflammatory; ophthalmological; antiallergic; antiasthmatic;  
XX nephroprotective; dermatological; antidiarrheic; vulnery; haemostatic;  
XX virucide; antibacterial; nootropic; neuroprotective; antiparkinsonian;  
XX anticonvulsant; analgesic; anorectic; metabolic; immunomodulator;  
XX proliferation; differentiation; cancer; apoptosis; leukaemia; arthritis;  
XX cardiovascular disease; diabetes mellitus; ulcerative colitis;  
XX Crohn's disease; glomerulonephritis; hepatocellular necrosis;  
XX homeostasis; meningitis; multiple sclerosis; pain; gene therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX FT 1.1479

XX FT /ftas= a /product= "Fucosyltransferase family member 32132"

XX MO200183721-A2.

XX 08-NOV-2001.

XX 27-APR-2001; 2001WO-US013805.  
XX 28-APR-2000; 2000US-0200604P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Meyers RA, Williamson M;

XX WPI; 2002-041492/05.

XX P-PSDB; AAM47905.

XX Novel isolated 32132 nucleic acid encoding 32132 polypeptide, a novel  
XX member of fucosyltransferase family, useful for treating atherosclerosis,  
XX multiple sclerosis, acute glomerulonephritis, Reynaud's disease, pain.

XX Claim 1; Page 105-106; 155pp; English.

XX The invention relates to an isolated 32132 nucleic acid (I), encoding a  
XX member of the fucosyltransferase family comprising: (a) nucleotide  
XX sequence having 80% identity to sequence S1 (ABO5333) or S3 (ABO5334);  
XX (b) comprising fragment of S1 or S3; (c) encoding a polypeptide (II)  
XX comprising a fully defined sequence S2 (AAM47905) of 492 amino acids or  
XX its fragment; or (d) encoding naturally occurring allelic variant of S2.  
XX (I) has cytosolic, antitumour, antiatherosclerotic, hypotensive,  
XX antiarrhythmic, vasotrophic, antidiabetic, antiallergic,  
XX immunosuppressive, antihypoid, antitumor, antiinflammatory,  
XX ophthalmological, antidiarrheic, vulnery, haemostatic, virucide,  
XX dermatological, nootropic, neuroprotective, antiparkinsonian,  
XX antibacterial, analgesic, anorectic, metabolic and immunomodulator  
XX activity. (II) and (III) are useful for treating proliferative and/or  
XX carcinogenic disorders of the colon, including adenoma and adenomas,  
XX breast, including epithelial hyperplasia and sclerosing adenosis, lung,  
XX including bronchogenic carcinoma and neuroendocrine tumour. The  
XX proliferative disorders treated also include haematopoietic neoplastic  
XX disorders such as acute promyeloid leukaemia, acute myelogenous  
XX leukaemia. (I) and (II) are also useful for treating cardiovascular  
XX diseases, disorders of blood vessels, immunological disorders such as  
XX diabetes mellitus, arthritis, ulcerative colitis, Crohn's disease,  
XX asthma, kidney disorders e.g., polycystic kidney disease, acute  
XX glomerulonephritis, colon disorders, disorders of liver such as disorders  
XX associated with accumulation of fibrous tissue, hepatocellular necrosis  
XX or injury induced by agents including processes which disturb  
XX homeostasis, disorders of brain such as intracranial hemorrhage,  
XX bacterial or viral meningitis, neurodegenerative disorder, glioma,  
XX multiple sclerosis, metabolic or pain disorders. (I) can be used in gene  
XX therapy

XX Sequence 1479 BP; 280 A; 453 C; 459 G; 287 T; 0 U; 0 Other;

XX Query Match 10.8%; Score 180.8; DB 6; Length 1479;

XX Best Local Similarity 51.8%; Pred. No. 2.4e-39;

XX Matches 489; Conservative 0; Mismatches 442; Indels 13; Gaps 3;

QY 440 GGTGGTCCCGCTACCGGGAGAGCTGGAGGTTAGGCCAATGTGAGAGAGAGCTTGT 499  
DB 252 GCTATTCCTCCCACTCCCGGAGAGCTGGAGAGCA-TCGAGTGTGGCGGGCGGTGG 310  
QY 500 TCTTACCATCAACCGGACCTACCTCATCATCATGACCAAGCAATTCCTCTTATAG 559  
DB 311 TGGCTCCCGGAGACCGCGAGCGCTGAGAGAGCTGGAGAGCGCGCGCTCTTCTACG 370  
QY 560 GTACTGACTTAAATAGATAGTACTTGTGCTCGGAAAGCCCATGATGATGAGCTG 619  
DB 371 GCACAGACTTCCGGCG 430  
QY 620 TTTTTCATGAGAGATCCCGGAAATTAATTAAGCTCTTCAATTAACGATGATGAT 679  
DB 431 TCTTCACAGAGAGTCCCGCTCAACAACTTCTTCTGAGGACACGCGCGCGATCGCC 490  
QY 680 TGTTCACATCACTGCGCAGTTCAGAGGAGATTCGCACTTGCACATCACTACCAAT 739  
DB 491 TCTTCAATCTTACCTCAGCTGAGGAGCTGAGATTCGCGCTGAGGAGGAG 550  
QY 740 TGGAGAGCAATTGAAGTCTGAGTCACTCGATACCTAGTTCCTTTCAGATCCAAAACA 799

Db 551 TGCCCGGAGACCGGCTACTCT---GGCGCCCGGCTGCTCCGCCCATGGAACCGCGGAGT 607  
Qy 800 AGCTTGAAGAAAAAGACTTCTCCGCTGGTGTATGATACAGTACAGACTGTGATCCACCATCAG 859  
Db 608 GGGCGCCCGCGGCTACGCGCCGCTCTCTATCTGTGAGTACACACTGCACGCGGCGGAG 667  
Qy 860 ACAAGGACAGCTATGTTGGAGCTGATGATTAACATGAGAGTGCATTCTATGGTAT 919  
Db 668 ACCGGACCGCTACGTCGGGAGCTCATGCGCAATCCCGGTAACTTCTACGGGAAT 727  
Qy 920 GTTTACGAAACAAAGACTTCCT-----CAGCAGCTGAAAAATCCAGCTCTATAG 970  
Db 728 GCCTGCAGATCGGGAGCTGCCCTACCGCGGCTACAGGACACACGCCGCAACCG 787  
Qy 971 ATGCCAGTGGCTTTTATAGATGATGACATTAATGATTTATCTTACTTTGAGATG 1030  
Db 788 AGATCCAGAGCTCTGGCTTTCTGTCCGCTAAATGATTCACCTTGGCCCTGGAATG 847  
Qy 1031 CAGTTGTGATGACTACATCACTGAGAAAGTTCTGAGGCGCACTGAACTGGGGGTATGC 1090  
Db 848 CCACTGTGATGACATGACATGACAGAAACCTGGCGTCCATGCACTGGGCGCTGTGC 907  
Qy 1091 CTGTATATTAACGATCCCCAGATCAGACTGGCTTCCAAATTAACAAAGTCTATTC 1150  
Db 908 CCGTGAACCGGCTTCCCTCTGTGAGGAGCTGATCCGAAACATCACTCCGCTATCC 967  
Qy 1151 TTGTATCAGAAATTTCTCACCCGAGGAACTGGCAAGTTATCATCAGACATGAGATTCTG 1210  
Db 968 TGAATGATGATTTTGAAGTCTCTCTCAGAGCTGGCAGATTTATGACTTTTGGACAAGA 1027  
Qy 1211 ATGACAGATTTATGAGGCGCTTATGAAATGAGAGCTGAAGGGTGAATCTCTTAACAGC 1270  
Db 1028 ATATGAGAGATTTATGAAATACCTGGCATACACAACTGGGGGCTATCCACCAAT 1087  
Qy 1271 GACTTCTGACAGCTCTCAGGAGAAAGGAGGATGCAAGCTCAACAGAGACAT 1330  
Db 1088 TTCTTTGATGATCTGAGAGCATCGGAGTGGGAGTAAATGATCTTTGCTGCTTACT 1147  
Qy 1331 ACATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1374  
Db 1148 ACCTCAACGCTTCGAGTGTTCGTCTGATGATGATGATGATGATGATGATGATGATGAT 1191

RESULT 11  
ABA05333  
ID ABA05333 standard; cDNA; 2557 BP.

AC ABA05333;

DT 26-FEB-2002 (first entry)

XX Human fucosyltransferase family member 32132 encoding cDNA.

XX Human; fucosyltransferase family member 32132; cytosolic; antitumor;  
XX antithrombotic; hypotensive; antiarrhythmic; vasotrophic;  
XX antidiabetic; antihypertensive; immunosuppressive; antihypertensive; antitumor;  
XX antidiabetic; antihypertensive; immunosuppressive; antihypertensive; antitumor;  
XX nephrotrophic; dermatological; antidiabetic; antihypertensive; antitumor;  
XX virulent; antibacterial; neurotrophic; neuroprotective; antiparkinsonian;  
XX anticonvulsant; analgesic; anorectic; metabolic; immunomodulator;  
XX proliferation; differentiation; cancer; apoptosis; leukaemia; arthritis;  
XX cardiovascular disease; diabetes mellitus; ulcerative colitis;  
XX Crohn's disease; glomerulonephritis; hepatocellular necrosis;  
XX homeostasis; meningitis; multiple sclerosis; pain; gene therapy; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 236..1714

FT /tag= a  
FT /product= "fucosyltransferase family member 32132"

PN WO200183721-A2.  
XX 08-NOV-2001.  
PD 27-APR-2001; 2001WO-US013805.  
PF 28-APR-2000; 2000US-0200604P.  
XX (MILL-) MILLENNIUM PHARM INC.  
PI Meyers RA, Williamson M;  
XX WPI: 2002-041492/05.  
DR P-PSDB; AAM47905.  
XX  
PT Novel isolated 32132 nucleic acid encoding 32132 polypeptide, a novel  
PT member of fucosyltransferase family, useful for treating atherosclerosis,  
PT multiple sclerosis, acute glomerulonephritis, Raynaud's disease, pain.  
PS Claim 1; Page 105-107; 125pp; English.  
XX  
CC The invention relates to an isolated 32132 nucleic acid (I), encoding a  
CC member of the fucosyltransferase family comprising: (a) nucleotide  
CC sequence having 80% identity to sequence S1 (ABA05333) or S3 (ABA05334);  
CC (b) comprising fragment of S1 or S3; (c) encoding a polypeptide (II)  
CC comprising a fully defined sequence S2 (AAM47905) of 492 amino acids or  
CC its fragment; or (d) encoding naturally occurring allelic variant of S2.  
CC (I) has cytosolic, antitumor, antithrombotic, hypotensive,  
CC antihypertensive, vasotrophic, antidiabetic, antihypertensive,  
CC immunosuppressive, antihypertensive, antitumor, antidiabetic,  
CC ophthalmological, antidiabetic, antihypertensive, nephrotrophic,  
CC dermatological, antidiabetic, antihypertensive, antitumor, antidiabetic,  
CC antibacterial, neurotrophic, neuroprotective, antiparkinsonian,  
CC anticonvulsant, analgesic, anorectic, metabolic and immunomodulator  
CC activity. (I) and (II) are useful for treating proliferative and/or  
CC differentiative disorders of the colon, including adenoma and colorectal  
CC carcinogenesis, the liver, including nodular hyperplasia and adenomas,  
CC breast, including epithelial hyperplasia and sclerosing adenosis, lung,  
CC including bronchogenic carcinoma and neuroendocrine tumour. The  
CC proliferative disorders treated also include haematopoietic neoplastic  
CC disorders such as acute promyeloid leukaemia, acute myelogenous  
CC leukaemia. (I) and (II) are also useful for treating cardiovascular  
CC diseases, disorders of blood vessels, immunological disorders such as  
CC diabetes mellitus, arthritis, ulcerative colitis, Crohn's disease,  
CC asthma, kidney disorders e.g., polycystic kidney disease, acute  
CC glomerulonephritis, colon disorders, disorders of liver such as disorders  
CC associated with accumulation of fibrous tissue, hepatocellular necrosis  
CC or injury induced by agents including processes which disturb  
CC homeostasis, disorders of brain such as intracranial hemorrhage,  
CC bacterial or viral meningitis, neurodegenerative disorder, glioma,  
CC multiple sclerosis, metabolic or pain disorders. (I) can be used in gene  
CC therapy  
XX  
SQ Sequence 2557 BP; 540 A; 690 C; 684 G; 641 T; 0 U; 2 Other;

Query Match 10.8%; Score 180.8; DB 6; Length 2557;  
Best Local Similarity 51.8%; Pred. No. 3e-39;  
Matches 489; Conservative 0; Mismatches 442; Indels 13; Gaps 3;

Qy 440 GGTGTCCTCCGCTGACGAGGAGAGCTGGAGTTAGCCAAATGAGCAGATGCTTTT 499  
Db 487 GCTATTCCTCCCACTTCCCGGAGATCTGGAGCCCA-TGAGTGTGCGCGCGGCTGG 545  
Qy 500 TCTTACCATCAACCGGACTTACCTCATATATACATGACCAAGATTCCTTTTATG 559  
Db 546 TGCGCTCCCGGAACCGCGAGCGCTGAGGACTCGCGGACGCGCGCTCTTTCTAG 605  
Qy 560 GTACTGACTTTAATCATGATAGCTTACCTTGCCTCGGAAAGCCATCATGATGGGCTG 619  
Db 606 GCACAGACTTCGCGGCGTGGCGCGCCCGCTGCGGCTGCGGACACAGACTGGGCGC 665  
Qy 620 TTTTTCATGAAGTCCCCGAAAAACATTTATAGTCTTTTATTAACCATGATTAATT 679

Db 666 TCCTCCAGAGAGAGTGGCCCTCAACAACTTCTTGCTGAGCCAGCGCCGGGATCCGCC 725  
QY 680 TGTTCACCTACCTGCGACAGTTGAGCAGGCAATTCCTCCACTGCACTACCAATACT 739  
Db 726 TCTTCATCTTACCTCCACCTTACCTCCGCACTCGGATTAACCCCTCTCGGTGAGTGGC 785  
QY 740 TGGAGAGCATTAAGTCCGTAAGTACTCCGATACCTTAAGTTCCCTTGAGTCCAAAACA 799  
Db 786 TGGCCGGGACCGCCCTATCT--GGCCGCCCGCGTCCCTCCGCCATGGAACGCGGAGT 842  
QY 800 AGCTTAGAAAAAGACTTCTCGCTGAGTGTATGTACAGTCAAGTGTGACCAACATCAG 859  
Db 843 GCGCGCGCGCGGTACGCGCGCGTGTCTATCTGCACTGCACTGCGAGCGTGCACGCG 902  
QY 860 AAGAGGACAGCATTTGTCGCGAGTGAAGTCACTTACATGAGTGCATTCCTTAAGTGAAT 919  
Db 903 ACCGGGACCGCTACCTGCGCGAGTCACTGCGCCACATCCCGTGAAGTCTTACGCGAAT 962  
QY 920 GTTTACGAACAAAGACTTCCT-----CAGCAGCTGAAAAATCCAGCCTCTATG 970  
Db 963 GCGTGAAGATGGGAGTGGCTTACCGCGCGGCTACAGGACAGCGACGCGCCACCG 1022  
QY 971 ATGCCGATGGCTTTATAGAGATCATGCAAGTATAGTTATCTTACGCTTTGAGATG 1030  
Db 1023 AGGATCCAGAGCTCTGGCTTCTTCTGTCGCCCTATTAAGTTCACTTGCCCTGAAAAATG 1082  
QY 1031 CAGTTGTGATGACTACATCACTGAGAGTCTGGAGGCCACTGAACTGGGGTATGTC 1090  
Db 1083 CCATCTGTACACATCACTACAGCAAAAACTGTGGCGTCCCATGCACTGGCGCTGTGC 1142  
QY 1091 CTGTATATTAGGATCCCCCAGACATCAGACTGCTTCCAAATACAAAAAGTCTATTTC 1150  
Db 1143 CCGTGTACCGCGGTCTCCCTCTGTGAGGAGCTGGATGCCGAAACATACCTCCGTATTC 1202  
QY 1151 TTGTATCAAGATTTCTCACCCCAGGAACTGGCAAGTTACATCAGACGATGATCTG 1210  
Db 1203 TGAATTGATTTTAACTCTCTCAAGAGCTGGCAGACTTATTAAGACTTTCTTGACAGA 1262  
QY 1211 ATGACAGATTGATGAGGCTATGTAGATGAAAGCTGAAGGATGAGATCTTAACCGAGC 1270  
Db 1263 ATGATGAGAGATATATGAATACCTGGCATACAGCAACCTCGGGGAGATCAACCAAT 1322  
QY 1271 GACTTGTGACAGCTCTCAGGGAACGGAATGGGAGTGCAGAGCTCAACAGGACAAATT 1330  
Db 1323 TTCTTCTGATAGTGTGAAGCATCGGAGTGGGAGTGAATGATCCCTTGTGCTTAAT 1382  
QY 1331 ACATCGATGATTTGATGATGATGTGTCACCAAGGTGTGGCT 1374  
Db 1383 AACTCAACGCTTCGAGTGTCTTCTGTGTGACTACGAACTGGCT 1426

RESULT 12  
ABLO7657  
ID ABLO7657 standard; cDNA; 1431 BP.  
XX  
AC ABLO7657;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 17453.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
XX  
pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW,  
XX  
DR MPI: 2001-656860/75.  
XX  
PT P-PSDB; ABB63554.  
XX  
PS New isolated nucleic acid detection reagent for detecting 1000 or more  
XX  
interactions.  
XX  
Claim 1; SEQ ID NO 17453; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB016176-AB016177), expressed DNA  
CC sequences (AB01840-AB01841) and the encoded proteins (AB057737-  
CC AB057742). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
XX  
SQ Sequence 1431 BP; 390 A; 341 C; 342 G; 358 T; 0 U; 0 Other;  
XX  
Query Match 10.8%; Score 179.8; DB 4; Length 1431;  
Best Local Similarity 54.7%; Pred. No. 4,4e-39;  
Matches 427; Conservative 0; Mismatches 342; Indels 12; Gaps 3;  
QY 476 GCCAATGTGAGAGAGATGCTTTCTTCAACATCAACCGAGACTACCTCATCATCA 535  
Db 272 GCGAGTGGGAAATTCATACCTTGGCGCATCAACCAAGGCGA-----GCCGCGGTCCGT 325  
QY 536 TGACCAAGCATTCCTCTTATGATGACTGACTTAAATATGATGATGCTTCTGCTC 555  
Db 326 GGGCCCGGGGGGCTCTGTTTACCGCTCAACATTAAGAGGAGAGACTTTCCTCCCTGCTC 385  
QY 596 GGAAGCCCATCATGACTGAGGCGCTTTTTCATGAGAAGTCCCGGAAAAAATATTAAGC 655  
Db 386 GGAAGAGACCAATCTGGCCCTTACTGATGAAGATCAACAGGAACACCCCTTTTG 445  
QY 656 TCTTTCATTAACCAAGTATTACTTGTCACTGACCTGCGAGTCAAGCAAGCTTCC 715  
Db 446 TGTCAATTAAGAGTCTTCTGCGACATTCATTCATCCAGTTTACGCGCTACAGCA 505  
QY 716 ACTGCGCATTAACCAATTACTTGAAGAGATGTAAGTCCGAAGTCACTCCGATACC 775  
Db 506 ACTTGCCGCTACATCAATGATGATCTGCCAGTGGCGAGCGCTTACTTCGAAGGATTA 565  
QY 776 TAGTTCCTTTCAGTCCAAAAACAAGCTTAGAAAAAG---ACTTGCTCCGCTGATG 832  
Db 566 ATGTTAGCTTTGAGGGGAGATCAAAAGTACGATACGTCATCCATCCATCGTGTGTC 625  
QY 833 TACAGTCACTGTGACCCACCATCAGACAGGACAGTATGTTGGAGCTGATGACT 892  
Db 626 TCCAAAGGAGCTGGACACGATGCTGAAAGGAGAGCTACGTAAGAACTGATGAAC 665  
QY 893 ACATCGAGTGCATTCCTATGATGATGATGATGATGATGATGATGATGATGATGATG 952  
Db 666 ATCTTCGATGATTTCTTACGAGAGCTGCTTACGGAATTAAGATCTCCAGAGAGGAAA 745  
QY 953 AAAATCCAGCTCTA--TGATGCCAGATGGCTTTTATAGATCATTCAGATATAGT 1009  
Db 746 AGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 805  
QY 1010 TTATGCTAGCTTTGAGATGCACTTTGATGATGATGATGATGATGATGATGATGATG 1069  
Db 806 TCATGATTCATTAAGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 865  
QY 1070 CACTGAAACTGGGGGATGATCCCTGTATATTAAGGATCCCGAGCATCAAGACTGCTTC 1129

DB 866 CTTTGAATAGGAGATTATTCATTTACTTTGGCTACCCACCTATAAAGCTGGGAC 925  
QY 1130 CAAGTAAACAAAGTCTATTCTTGTATCGAATTTTTCACCCCGGAACTGGCACTT 1189  
DB 926 CGAATATAATATCGAATTTTGTACAGACTTTCAAAATCTCAGGCACTGTAGAGT 985  
QY 1190 ACATCAGACGACTGATTCGTATGATGAGGCGCTATGAGATGAGAGCTGA 1249  
DB 986 ATCTCATATAGTTGGGGGATACAAAAGCTGTCAATTCTGACCCCAACAACTAA 1045  
QY 1250 A 1250  
DB 1046 A 1046

## RESULT 13

ACC62479 ID ACC62479 standard, cDNA, 1804 BP.

ACC62479;

23-JUN-2003 (first entry)

Human secreted protein #5 coding sequence SEQ ID 15.

Human, cardiac; hypotensive; hypertensive; cytosolic; antiinflammatory;  
antiasthmatic; osteopathic; antithyroid; gynecological; antibacterial;  
fungicide; virucide; antiparasitic; protozoicide; antianemic;  
antiarthritic; immunosuppressive; antirheumatic; antipsoriatic;  
haemostatic; antidiabetic; antiallergic; antigout; antiarrhythmic;  
antipainful; nephrotoxic; neuroleptic; neuroprotective; noctropic;  
antiparkinsonian; tranquiliser; anorectic; antituber; hepatotropic;  
dermatological; gene therapy; neoplastic disease; blood disorder;  
cardiovascular disorder; respiratory disorder; cancer;  
endocrine disorder; reproductive system disorder; infection; arthritis;  
asthma; autoimmune disease; immune deficiency; psoriasis;  
diabetes mellitus; allergy; Paget's disease; gout; osteoporosis;  
renal disorder; Alzheimer's disease; Parkinson's disease; schizophrenia;  
attention deficit disorder; obsessive compulsive disorder; wound healing;  
gene; ss.

Homo sapiens.

MO200299066-A2.

12-DEC-2002.

05-JUN-2002; 2002MO-US017699.

06-JUN-2001; 2001US-0295869P.

11-JUL-2001; 2001US-0304121P.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Bell A, Birse CE, Komatsoulis G, Choi GH, Olsen H;

N1 J, Baker KP;

WPI; 2003-140609/13.

Newel human secreted polypeptides and polynucleotides for diagnosing and

cardiovascular, renal, proliferative disorders and cancerous diseases.

Claim 1, Page 539-540; 605bp; English.

The present invention relates to novel human secreted proteins (ABR54278-  
ABR54331) and their coding sequences (ACC62475-ACC62528). The proteins  
and their coding sequences are useful for treating, preventing,  
diagnosing and/or prognosing neoplastic diseases, blood disorders,  
cardiovascular disorders (e.g. cardiomyopathy, hypertension,  
hypotension), respiratory disorders (e.g. lung cancer, pneumonia,

bronchiolitis, asthma), endocrine disorders (e.g. Addison's disease,  
Cushing's syndrome, hyperpituitarism, pituitary dwarfism, acromegaly,  
thyroiditis), reproductive system disorders (e.g. premenstrual syndrome,  
polycystic ovary syndrome), infectious diseases caused by bacteria,  
fungal, viral, parasitic, protozoal, and/or blood-related disorders and  
infections, leukopaenia, leukemias, arthritis, asthma, autoimmune  
diseases, rheumatoid arthritis, immune deficiency, psoriasis,  
haemophilia, diabetes mellitus, allergies, and bone cancers, Paget's  
disease, gout, osteoporosis, arrhythmia, angina, prostate cancer, renal  
disorders, urolithiasis, Alzheimer's disease, Parkinson's disease,  
schizophrenia, attention deficit disorder, obsessive compulsive  
disease, pneumonia, obesity, golfer's ulcerative colitis, hepatitis. They are also  
useful for stimulating epithelial cell proliferation and basal  
keratinocytes for wound healing, and to stimulate hair follicle  
production and healing of dermal wounds

Sequence 1804 BP; 405 A; 474 C; 457 G; 465 T; 0 U; 3 Other;

Query Match 9.7%; Score 162.6; DB 7; Length 1804;

Best Local Similarity 52.5%; Pred. No. 2.9e-34;

Matches 410; Conservative 0; Mismatches 359; Indels 12; Gaps 2;

QY 603 CCATCATGACCTGGGCTGTTTTCATGGAAGTCCCGCAAAAACATATATAGCTCTTCA 662  
DB 158 CCTTGCGCGCTGGGCGCTCTCCACGAGATCCCGCTCAACACTCTTCTGTAGCCA 217  
QY 663 TAAACCACTGATTAACCTTGTTCACACTACGACGCTTACGAGGAGTCCACTTGGC 722  
DB 218 CGCGCCGGGCAATCCGCGCTTCAATCTTACTCCACTTCAGTGCACATGATTAACC 277  
QY 723 ACTAATCCCAATTAATCTTGAGAGACATGAAATCTGAATCACTCCGATTAATTC 782  
DB 278 GCTGTGCTGAGTGGCTGCGCGGACCGCTATCTGCG--CGCGGTGCTTCCGCC 334  
QY 783 TTGAGAGTCCAAAACAAGCTTAAGAAAAGACTTGCCTCGCTGATATAGTACAG 842  
DB 335 CATGGAAGCGCGGAGTGGCGCGCGCGCGGTACGCGCGCTGATATGAGTACAG 394  
QY 843 CTGTGACCCACCATAGACAGAGGACAGCTATGTTGCGAGTGAATCACTTACATCA 902  
DB 395 CTGCGACGTGCGAGGAGCGCGGACCGCTACGTGCGGAGCTATGCGCACATCCGG 454  
QY 903 CGATTCCTATGATGATTTATAGCAAAACAAGACTTCCCT-----CAGACGTGAA 953  
DB 455 AGACTCTTAAGGAAATCCGCAAAATCGGAGCTCTTACCGCGGCGTACAGAGAC 514  
QY 954 AATTCAGCTCTATGATGATGCGGATGCGATGCTTTATAGGATCATTCACAGTAT 1013  
DB 515 AGCCACGGCCACACCGGAGATCCAGAGCTTGGCTTTCTTGCCCGCTATAGTCCA 574  
QY 1014 CCTACTTTGAGATGATGATTTGATGATGATGATGATGATGATGATGATGATGAT 1073  
DB 575 CTGGCCCTGGAATATGATCTTTAAGACTATATACGAAAAATGTTGGCTCCAT 634  
QY 1074 GAAACTGGGGTATGCTGATATTAAGATCCCGACGATCAGACACTGGCTTCAAG 1133  
DB 635 GCACCTGGCGCTGCGCGGTATCCGCGGCTTCCCTGTGAGGAGACTGATGCGAA 694  
QY 1134 TAAACAAAGTCTATCTTCTATCAGATTTTTCACCCGAGGAGACTGCAAT 1193  
DB 695 CATACATCTCTATCTCTGATGATGATTTGATGCTTCTTCAAGAGTGGCAAGTTTAT 754  
QY 1194 CAGACGATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1253  
DB 755 TGACTTTTGCAAGAAATGATGAGAGTATATGAAATACCTGGCATACAGCACTGG 814  
QY 1254 TGAGATCTCTAACACGAGCTTCTGACAGCTTCAAGGAAACGAAATGGGAGATG 1313  
DB 815 GGGCATTCACCAACAAATTTCTTCTGATGATGATGATGATGATGATGATGATG 874  
QY 1314 CGTCAACAGAGCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATG 1373  
DB 875 TCCTTGTGCTTAATCTTCAACGCGCTTCAAGTCTTGTCTGTGATCAAGACTGGC 934

QY 1374 T 1374  
 Db 935 T 935

RESULT 14  
 ACC62499  
 ID ACC62499 standard; cDNA; 1812 BP.  
 XX ACC62499;  
 XX 23-JUN-2003 (first entry)  
 DE Human secreted protein #25 coding sequence SEQ ID 35.

Human; cardiac; hypotensive; hypertensive; cytosolic; antiinflammatory;  
 antiaesthetic; osteopathic; antihypertoid; gynecological; antibacterial;  
 fungicide; virucide; antiparasitic; procoagulant; antianaemic;  
 antiarthritic; immunosuppressive; antineumatic; antiporiatic;  
 haemostatic; antidiabetic; antiallergic; antigenic; antiarrhythmic;  
 antiangiogenic; nephrotoxic; neuroleptic; neuroprotective; nootropic;  
 antiparkinsonian; tranquiliser; anorectic; antituber; hepatotropic;  
 dermatological; gene therapy; neoplastic disease; blood disorder;  
 cardiovascular disorder; reproductive system disorder; cancer;  
 endocrine disorder; reproductive system disorder; infection; arthritis;  
 asthma; autoimmune disease; immune deficiency; psoriasis;  
 diabetes mellitus; allergy; Paget's disease; gout; osteoporosis;  
 renal disorder; Alzheimer's disease; Parkinson's disease; schizophrenia;  
 attention deficit disorder; obsessive compulsive disorder; wound healing;  
 gene, ss.

XX Homo sapiens.  
 XX MO200299066-A2.  
 XX 12-DEC-2002.  
 XX 05-JUN-2002; 2002MO-US017699.  
 XX 06-JUN-2001; 2001US-0295869P.  
 XX 11-JUL-2001; 2001US-0304121P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Ruben SM, Bell A, Birse CE, Komatsoulis G, Choi GH, Olsen H;  
 XX Ni J, Baker KP;  
 XX MPI: 2003-140609/13.  
 XX P-PSDB; ABR54302.  
 XX Novel human secreted polypeptides and polynucleotides for diagnosing and  
 XX treating neural, immune system, muscular, reproductive, gastrointestinal,  
 XX cardiovascular, renal, proliferative disorders and cancerous diseases.  
 XX Claim 1; Page 551; 605pp; English.

XX The present invention relates to novel human secreted proteins (ABR54278-  
 XX ABR54331) and their coding sequences (ACC62475-ACC62528). The proteins  
 XX and their coding sequences are useful for treating, preventing,  
 XX diagnosing and/or prognosing neoplastic diseases, blood disorders,  
 XX cardiovascular disorders (e.g. cardiomyopathy, hypertension,  
 XX hypotension), respiratory disorders (e.g. lung cancer, pneumonia,  
 XX bronchitis, asthma), endocrine disorders (e.g. Addison's disease,  
 XX Cushing's syndrome, hyperpituitarism, pituitary dwarfism, acromegaly,  
 XX thyroiditis), reproductive system disorders (e.g. premenstrual syndrome,  
 XX polycystic ovary syndrome), infectious diseases caused by bacteria,  
 XX fungal, viral, parasitic, protozoal, and/or blood-related disorders and  
 XX infections, leukaemia, leukaemias, arthritis, asthma, autoimmune  
 XX diseases, rheumatoid arthritis, immune deficiency, psoriasis,  
 XX haemophilia, diabetes mellitus, allergies, and bone cancers, Paget's  
 XX disease, gout, osteoporosis, arthralgia, angina, prostate cancer, renal  
 XX disorders, urolithiasis, Alzheimer's disease, Parkinson's disease,

CC schizophrenia, attention deficit disorder, obsessive compulsive  
 CC pneumonia, obesity, goiter, ulcerative colitis, hepatitis. They are also  
 CC useful for stimulating epithelial cell proliferation and basal  
 CC keratinocytes for wound healing, and to stimulate hair follicle  
 CC production and healing of dermal wounds  
 CC XX  
 SQ Sequence 1812 BP; 431 A; 480 C; 461 G; 440 T; 0 U; 0 Other;

Query Match 9.7%; Score 162.6; DB 7; Length 1812;  
 Best Local Similarity 52.5%; Pred. No. 2.9e-34;  
 Matches 410; Conservative 0; Mismatches 359; Indels 12; Gaps 2;

QY 603 CCATCATGCTGGGCTGTTTTCATGAGAGTCCCGAAACAAATTAATTAAGCTCTTCA 662  
 Db CCTCGGGGCTGGGCGGCTCCACGAGAGTGGCCCTCAACAACTTCTGCTGAGCCA 235

QY 663 TAAACAGTATTTCTTTGTTCACTACCTGCCAGTTAGAGGATTTCCACTTGC 722  
 Db CGGCCCGGAGCTCCGCTCTTCACTTACCTCACCCTTAGTGGCAGCTGGATTACC 295

QY 723 ACTAATCACTACATATGAGAGCATTTGAAGTCTGAATCACTCCGATACCTAGTTC 782  
 Db GCTGCTGCTGAGTGGCTGCCCGGAGCCCTATCTGCG---CGCCGGTGGCTCCGCC 352

QY 783 TTTGAGTCAAAAAACAAGCTTAGAAAAAGACTTCTCCGCTGGTGTATGATGACGA 842  
 Db CATGGAACGCCCGAGTGGCGCCCGGCTACGCCCGCTGCTATCTGCACTGACACA 412

QY 843 CTGTGACCCACATCAGACAGGACAGCTATGTTGGAGCTGATGATCTACATCGAGT 902  
 Db CTGCACCTGCGACGGACCGGACCGCTACGTCGCGAGCTCATGGCCACATCCCGT 472

QY 903 CGATTCTANTGATGATGTTTACGAAACAAAGACTTCCCT-----CAGACGTGAA 953  
 Db AGACTCTACGGAAGATGCTGAGATTCGGAGCTCTACCGCGGCTACAGAGAC 532

QY 954 AATTCAGCTCTTATGATGATGCGATGCTTTTATAGATCATGACAGATTAAGTTAT 1013  
 Db AGCAGGGCACACCGAGGATCCAGAGCTCTTGCTTTTGTCTCCCTATTAAGTTCCA 552

QY 1014 CCTAGCTTTAGAGATGAGATTTGATGATGATGATGATGATGATGATGATGATGAT 1073  
 Db CTGGCCCTGGAATAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 652

QY 1074 GAACTGGGGGTAGTCCCTGTATATTTACGATCCCCAGATCAAGACTGGCTTCAAG 1133  
 Db GCACCTGGGGCTGTGCGCGGTGTACCGGGTTCCTCTGTGAGGACCTGATGCGGA 712

QY 1134 TAACAAAGTCTATTTCTTGTATCAGAAATTTTTCACCCAGGGAAGTGGCAATTCAT 1193  
 Db CATATCTCCGTCATCTGATGATGATTTGATGATGATGATGATGATGATGATGAT 772

QY 1194 CAGACGATGATTTCTGATGACAGATTGTAGGCTTATGATGATGATGATGATGATG 1253  
 Db TGACTTTCTGACAAAGATGATGAGATGATGATGATGATGATGATGATGATGATG 832

QY 1254 TGAGATCTCTAACAGGAGCTTCTGACAGCTCTCAGGGAACGGAATGGGAGTGCAGA 1313  
 Db GGGCATCCCAACCAATTTCTTCTGATGATGATGATGATGATGATGATGATGATG 892

QY 1314 GGTCAACAGGACAAATTAATGATGATGATGATGATGATGATGATGATGATGATG 1373  
 Db TCCCTTGTGCTTAATCACTCAACGAGGCTTGAATGTTTGTCTGATCAAGAACTGGC 952

QY 1374 T 1374  
 Db 953 T 953

RESULT 15  
 ACC62500  
 ID ACC62500 standard; cDNA; 1804 BP.

AC ACC62500;  
 XX 23-JUN-2003 (first entry)  
 XX Human secreted protein #26 coding sequence SEQ ID 36.  
 DE  
 XX Human: cardiac; hypotensive; hypertensive; cytostatic; antiinflammatory;  
 KM antiasthmatic; osteoplastic; antihypertoid; gynecological; antibacterial;  
 KM fungicide; virucide; antiparasitic; protozoacide; antianemic;  
 KM antiallergic; immunosuppressive; antirheumatic; antipsoriatic;  
 KM haemostatic; antidiabetic; antiallergic; antigout; antiarhythmic;  
 KM antiangiinal; nephrotoxic; neuroleptic; neuroprotective; noctropic;  
 KM antiparkinsonian; tranquiliser; anorectic; antitumor; hepatotropic;  
 KM dermatological; gene therapy; neoplastic disease; blood disorder;  
 KM cardiovascular disorder; respiratory disorder; cancer;  
 KM endocrine disorder; reproductive system disorder; infection; arthritis;  
 KM asthma; autoimmune disease; immune deficiency; psoriasis;  
 KM diabetes mellitus; allergy; Paget's disease; gout; osteoporosis;  
 KM renal disorder; Alzheimer's disease; Parkinson's disease; schizophrenia;  
 KM attention deficit disorder; obsessive compulsive disorder; wound healing;  
 KM gene; ss.  
 XX  
 OS Homo sapiens.  
 PN WO200299066-A2.  
 PD 12-DEC-2002.  
 XX 05-JUN-2002; 2002MO-US017699.  
 PE  
 XX 06-JUN-2001; 2001US-0295869P.  
 PR 11-JUL-2001; 2001US-0304121P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Ruben SM, Bell A, Birse CE, Komatsoulis G, Choi GH, Olsen H;  
 PI Ni J, Baker KP;  
 XX WPI: 2003-140609/13.  
 DR P-PSDB; ABR54303.  
 XX  
 PT Novel human secreted polypeptides and polynucleotides for diagnosing and  
 PT treating neural, immune system, muscular, reproductive, gastrointestinal,  
 PT cardiovascular, renal, proliferative disorders and cancerous diseases.  
 PT  
 XX Claim 1; Page 551-552; 605pp; English.  
 PS  
 XX The present invention relates to novel human secreted proteins (ABR54278-  
 CC ABR54331) and their coding sequences (ACC62475-ACC62528). The proteins  
 CC and their coding sequences are useful for treating, preventing,  
 CC diagnosing and/or prognosing neoplastic diseases, blood disorders,  
 CC cardiovascular disorders (e.g. cardiomyopathy, hypertension,  
 CC hypotension), respiratory disorders (e.g. lung cancer, pneumonia,  
 CC bronchiolitis, asthma), endocrine disorders (e.g. Addison's disease,  
 CC Cushing's syndrome, hyperpituitarism, pituitary dwarfism, acromegaly,  
 CC thyroiditis), reproductive system disorders (e.g. premenstrual syndrome,  
 CC polycystic ovary syndrome), infectious diseases caused by bacteria,  
 CC fungal, viral, parasitic, protozoal, and/or blood-related disorders and  
 CC infections, leukopenia, leukemias, arthritis, asthma, autoimmune  
 CC diseases, rheumatoid arthritis, immune deficiency, psoriasis,  
 CC haemophilia, diabetes mellitus, allergies, and bone cancers, Paget's  
 CC disease, gout, osteoporosis, arrhythmia, angina, prostate cancer, renal  
 CC disorders, urolithiasis, Alzheimer's disease, Parkinson's disease,  
 CC schizophrenia, attention deficit disorder, obsessive compulsive  
 CC pneumonia, obesity, goiter, ulcerative colitis, hepatitis. They are also  
 CC useful for stimulating epithelial cell proliferation and basal  
 CC keratinocytes for wound healing, and to stimulate hair follicle  
 CC production and healing of dermal wounds  
 XX  
 SQ Sequence 1804 BP; 405 A; 473 C; 456 G; 465 T; 0 U; 5 Other;

Matches 409; Conservative 1; Mismatches 359; Indels 12; Gaps 2;  
 QY 603 CCATGATGACTGGGCTGTTTTCATGAAAGTCCCGGAAAAAATTATTAAGCTTTTCA 662  
 Db |||||  
 Db 158 CCTCGGGGCTGGGCGCTCTCCACGAGAGTCCCGCTCAACAACATCTTGGTGAGCCA 217  
 QY 663 TAAACAGGATATTAAGCTTGTTCACATACAGCCAGCTGAGAGGATCCACATGCGC 722  
 Db |||||  
 Db 218 CGGCGGCGATCCGCTCTTCAATCTTCACTCCACCTTGATGCGCACTCGATTACCC 277  
 QY 723 ACTAATCAACCAATCTTGAAGACATTGAAGTCTGAAGTCACTCGATTAAGTTC 782  
 Db |||||  
 Db 278 GCTGTGCTGAGTGGTGGCCCGGAGCCGCTATCTGGC---CGCCGGTGCTCGCGC 334  
 QY 783 TTTGAGTCCAAAACAGCTTAAAGAAAAGCTTGCTCGGTGGTGTATGAGTGA 842  
 Db |||||  
 Db 335 CATGAAAGCGGGGAGTGGCCCGCGGCTACCGCGCTGCTTATCTGCAATCA 394  
 QY 843 CTGTGACCCACCATCAGACAGAGACAGTATGTCGAGCTGATGACTTACATCGAGT 902  
 Db |||||  
 Db 395 CTGGACGTCGAGCGGACCGGACCGGTAAGTGGAGCTATGCGCACATCCGGT 454  
 QY 903 CGATTCCTATGTTGATTTTACGAAACAAAGACCTCCCT-----CAGACGTGMA 953  
 Db |||||  
 Db 455 AGACTCTTACGGGAAATGCTCGAGAAATCGGAGCTGCTTACCGCGGGGTACAGAC 514  
 QY 954 AATCCAGCTCTATGATGATGCGGAGTGGCTTTATAGATCATTTGACAGTATAGTTAT 1013  
 Db |||||  
 Db 515 AGCCAGGCGCACACAGAGATCCAGAGCTTTGGCTTTCTTGCCCGCTATAAGTTCCA 574  
 QY 1014 CTTAGCTTTGAGAAATGAGTGTGTGATGACTATCATCATGAGAAAGTTCTGAGGCCACT 1073  
 Db |||||  
 Db 575 CTGGCCCTGGAATAATGCATCTGTAACTATCATCAGAAAACATGCGGCTCCCAT 634  
 QY 1074 GAACTGGGGGAGTCCCTGTATTTAGGATCCCGCAGATCAGACTGGCTTCAAG 1133  
 Db |||||  
 Db 635 GCACTGGGCGGTGGCCGCTGACCGGGTCTTCCCTGTGTAGGAGCTGATGCGGAA 694  
 QY 1134 TACAAAAGTCTATTTCTGTATCAGAAATTTCTACCCAGGAACTGGCAATTACAT 1193  
 Db |||||  
 Db 695 CATCATCCCGATCCGATGATTGATGATTTGAGTCTCTCAGAGCTGCAAGATTAT 754  
 QY 1194 CAGAGACTGATTTGTATGACATTTGTATGAGGCTATGTAAATGAAAGTGAAGG 1253  
 Db |||||  
 Db 755 TCACCTTGTGACAAAGATGATGAGATATGAAATACCTGSCATACAGCAACCTGG 814  
 QY 1254 TGAGATCTCTAACCGACGACTTTCAGAGCTCTCAGGGAACGGAATGGGAGTGAAGA 1313  
 Db |||||  
 Db 815 GGGCATACCAACCAATTTCTTGATAGTCTGAACCAACGGAGTGGGAGATGATGA 874  
 QY 1314 CCTCAACGAGCAATTCATGATGATGATTTGATGATGATGATGATGATGATGATG 1373  
 Db |||||  
 Db 875 TCCTTTGTGCTACTACCTTAACCTTAACGCTTGAAGTGTGCTGTGATCAACACTGC 934  
 QY 1374 T 1374  
 Db 935 T 935

Search completed: September 12, 2004, 22:22:31  
 Job time : 477.373 secs





Page 2

TELECOMMUNICATION INFORMATION  
TELEPHONE: (703) 836-9300

1118 CAGACTGCGCTTCCAGTAACAAGAAGTGTATTTCTGTATCAGAAATTTCTCACCCCAAGG 1177



```

STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476.062A
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216.081
FILING DATE: 21-MAR-1994
APPLICATION NUMBER: 07/637.830
FILING DATE: 04-JAN-1991
APPLICATION NUMBER: 07/539.842
FILING DATE: 18-JUN-1990
APPLICATION NUMBER: 07/212.573
FILING DATE: 28-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00786/068003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 4704 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-476-062A-52

Query Match
Best Local Similarity 62.2%; Score 38.8; DB 2; Length 4704;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1572 GATCAAACTTTTCATCTCAAGATTGGGCGCTGATTCAGAGCTGATTTCAAAA 1631
DB 4582 GCACAAAACGATGATCTACCGCTCGGAAATATCTGAAGGCTTAATAAAAAA 4641
QY 1632 ATGATCAGATGAACAGAAAAA 1669
DB 4642 AAGCCTTCTGTGAAAAA 4679

RESULT 9
US-09-023-655-1358
Sequence 1358, Application US/09023655
Patent No. 6607879
GENERAL INFORMATION:
APPLICANT: Cocks, Benjamin G.
APPLICANT: Susan G. Stuart
APPLICANT: Jeffrey J. Seihamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
TITLE OF INVENTION: EXPRESSION
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,655
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1358:
SEQUENCE CHARACTERISTICS:
LENGTH: 4704 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g35175
US-09-023-655-1358

Query Match
Best Local Similarity 62.2%; Score 38.8; DB 4; Length 4704;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1572 GATCAAACTTTTCATCTCAAGATTGGGCGCTGATTCAGAGCTGATTTCAAAA 1631
DB 4582 GCACAAAACGATGATCTACCGCTCGGAAATATCTGAAGGCTTAATAAAAAA 4641
QY 1632 ATGATCAGATGAACAGAAAAA 1669
DB 4642 AAGCCTTCTGTGAAAAA 4679

RESULT 10
US-09-621-976-16983
Sequence 16983, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 16983
LENGTH: 277
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-16983

Query Match
Best Local Similarity 65.1%; Score 38; DB 4; Length 277;
Matches 56; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 1584 TTCATCTCAAGATTGGGCGCTAGTATTCAGAGCTGATTTCAAAAATGATCAGATG 1643
DB 167 TTCTTCTCTCAGATTTGGGAGTGAAGAGAGACCGCTCGAATATATAAAAAAG 226
QY 1644 AAACGAAAAA 1669
DB 227 AAAAAA 252

RESULT 11

```

US-09-621-976-16052  
; Sequence 16052, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; NUMBER OF SEQ ID NOS: 2000-07-21  
; SOFTWARE: Patent.pm  
; SEQ ID NO 16052  
; LENGTH: 329  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc.feature  
; LOCATION: 279,281..282  
; OTHER INFORMATION: n=a, g, c or t  
US-09-621-976-16052

Query Match 2.3%; Score 38; DB 4; Length 329;  
Best Local Similarity 51.7%; Pred. No. 0.057;  
Matches 77; Conservative 2; Mismatches 70; Indels 0; Gaps 0;

QY 1521 CTTGACATCCAGAGAGCCGACCTAGGCTGGCTGATGATCAAA 1580  
DB 178 CTGTACCAGAAATATAATATATCGATAGAAAATATGAAATGAAA 237  
QY 1581 CTTTCATCTCAGAGTTTGGGGCTTAGTTCAGGACTGATTTCAAAAATGATCAGA 1640  
DB 238 AAATTTAAAAAACCAACCAAGCTCATCTATGAAAAAANNNNNAAAAA 297  
QY 1641 ATGAACAGAAAAA 1669  
DB 238 AAAAAAAAAAAAAAAAAAAAAA 326

RESULT 12  
US-09-800-729-35  
; Sequence 35, Application US/09800729  
; Patent No. 6605592  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 32 Human secreted proteins  
; FILE REFERENCE: P2044P1  
; CURRENT APPLICATION NUMBER: US/09/800,729  
; CURRENT FILING DATE: 2001-03-08  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: PCT/US00/26013  
; PRIOR FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 217  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 35  
; LENGTH: 6065  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: SITE  
; LOCATION: (6035)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (6037)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (6038)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-800-729-35

Query Match 2.3%; Score 38; DB 4; Length 6065;

Best Local Similarity 57.6%; Pred. No. 0.47;  
Matches 68; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1552 CTAAGTGCTGCTGTGADGGAATCAAACTTTTCATCTCAAGAGTTTGGGCTTAGTA 1611  
DB 5876 CTTAGGGGGAACATTGTAAAGAAAACAAAAGTCCAGATGATGTAGTAAATAAA 5935  
QY 1612 TTCAGACGATTTCAAAAATGATCGAATGAACAGAAAAA 1669  
DB 5936 GTTGAAGATTCTTA 5993

RESULT 13  
US-09-443-041A-27  
; Sequence 27, Application US/09443041A  
; Patent No. 6465717  
; GENERAL INFORMATION:  
; APPLICANT: Famodu, Omolayo O.  
; APPLICANT: Orozco, Buddy  
; APPLICANT: Rafalski, Antoni  
; APPLICANT: Shen, Jennie  
; TITLE OF INVENTION: Sterol Metabolism Enzymes  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/09/443,041A  
; CURRENT FILING DATE: 1999-11-18  
; PRIOR APPLICATION NUMBER: 60/109,283  
; PRIOR FILING DATE: 1998-11-20  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 27  
; LENGTH: 1447  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-443-041A-27

Query Match 2.3%; Score 37.8; DB 4; Length 1447;  
Best Local Similarity 58.4%; Pred. No. 0.19; 47; Indels 0; Gaps 0;  
Matches 66; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1557 GTGGCTGTTGATGAGATCAAACTTTTCATCTCAAGAGTTTGGGCTTAGTATCA 1616  
DB 1263 GTGTTTGTTGTAAGAGAAAAAGATGGGATGGAATAGGTGAGAGAGAAAA 1322  
QY 1617 GGAAGTATTTCAAAAATGATCAGATGAACAGAAAAA 1669  
DB 1323 AAAAAAAAAAAAAAAAAAAAAA 1375

RESULT 14  
US-08-021-608D-9  
; Sequence 9, Application US/08021608D  
; Patent No. 5580760  
; GENERAL INFORMATION:  
; APPLICANT: LEVENS, DAVID L., DUNCAN,  
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.  
; TITLE OF INVENTION: NOVEL FUSE BINDING  
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESSES:  
; ADDRESS: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/021,608D  
; FILING DATE: 22-FEB-1993

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: WILLIAM S. FEILER  
REGISTRATION NUMBER: 26,728  
REFERENCE/DOCKET NUMBER: 2026-4063  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2381  
TYPE: Nucleic acid  
STRANDEDNESS: Double  
TOPOLOGY: Unknown  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: No  
ANTI-SENSE: YES  
ORIGINAL SOURCE:  
ORGANISM: Human  
CELL LINE: HL60  
FEATURE:  
OTHER INFORMATION: 470 bp variable  
US-08-021-609D-9  
region where R is A or G.  
Query Match 2.3%; Score 37.8; DB 1; Length 2381;  
Best Local Similarity 61.9%; Pred. No. 0.28;  
Matches 60; Conservative 0; Mismatches 37; Indels 0; Gaps 0;  
Qy 1573 AATCAAACTTTTCATCTCAAGAGTTTGGGGCTAGTATTCAGAGCTGATTTCAAAA 1632  
Db 2260 AATGTAATCTTTCACCTTTCACAAATGCCTGTTTGTGCTTACATAAATGATATGAACC 2319  
Qy 1633 TGATCAGATGAACGAAAAA 1669  
Db 2320 TCAAAAAAAAAAAAAAAAAAAAAA 2356  
RESULT 15  
US-08-726-160-9  
Sequence 9, Application US/08726160  
Patent No. 5734016  
GENERAL INFORMATION:  
APPLICANT: LEVENS, DAVID L., DUNCAN,  
APPLICANT: ROBERT C., AND AVIGAN, MARK I.  
TITLE OF INVENTION: NOVEL FUSE BINDING  
TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/726,160  
FILING DATE: 04-OCT-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/021,608  
FILING DATE: 22-FEB-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: WILLIAM S. FEILER  
REGISTRATION NUMBER: 26,728  
REFERENCE/DOCKET NUMBER: 2026-4063US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2381  
TYPE: Nucleic acid  
STRANDEDNESS: Double  
TOPOLOGY: Unknown  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: No  
ANTI-SENSE: YES  
ORIGINAL SOURCE:  
ORGANISM: Human  
CELL LINE: HL60  
FEATURE:  
OTHER INFORMATION: 470 bp variable  
US-08-726-160-9  
region where R is A or G.  
Query Match 2.3%; Score 37.8; DB 1; Length 2381;  
Best Local Similarity 61.9%; Pred. No. 0.28;  
Matches 60; Conservative 0; Mismatches 37; Indels 0; Gaps 0;  
Qy 1573 AATCAAACTTTTCATCTCAAGAGTTTGGGGCTAGTATTCAGAGCTGATTTCAAAA 1632  
Db 2260 AATGTAATCTTTCACCTTTCACAAATGCCTGTTTGTGCTTACATAAATGATATGAACC 2319  
Qy 1633 TGATCAGATGAACGAAAAA 1669  
Db 2320 TCAAAAAAAAAAAAAAAAAAAAAA 2356  
Search completed: September 13, 2004, 02:21:04  
Job time : 92.1818 secs

| Result No. | Score  | Query Match | Length | DB | ID                  | Description         |
|------------|--------|-------------|--------|----|---------------------|---------------------|
| 1          | 1669   | 100.0       | 1669   | 14 | US-10-080-960-1     | Sequence 1, Appl    |
| 2          | 1622   | 97.2        | 2937   | 10 | US-09-814-333-30757 | Sequence 20757, A   |
| 3          | 1517.2 | 90.9        | 1942   | 9  | US-09-731-872-43    | Sequence 43, Appl   |
| 4          | 1517.2 | 90.9        | 1942   | 10 | US-09-876-997-43    | Sequence 43, Appl   |
| 5          | 1461   | 87.5        | 1461   | 14 | US-10-080-960-3     | Sequence 3, Appl    |
| 6          | 581.2  | 34.8        | 587    | 11 | US-09-864-408A-7105 | Sequence 7105, Appl |
| 7          | 453    | 27.1        | 530    | 10 | US-09-814-333-16082 | Sequence 16082, A   |
| 8          | 450.6  | 27.0        | 455    | 10 | US-09-814-333-3379  | Sequence 3379, Appl |
| 9          | 450.6  | 27.0        | 455    | 10 | US-09-814-333-3698  | Sequence 3698, Appl |
| 10         | 323.8  | 19.4        | 378    | 9  | US-09-764-877-187   | Sequence 187, Appl  |
| 11         | 323.8  | 19.4        | 378    | 16 | US-10-242-515-187   | Sequence 187, Appl  |
| 12         | 298    | 17.9        | 1550   | 17 | US-10-467-595-50    | Sequence 50, Appl   |
| 13         | 180.8  | 10.8        | 1479   | 10 | US-09-844-948-3     | Sequence 3, Appl    |
| 14         | 180.8  | 10.8        | 1479   | 13 | US-10-184-648-21    | Sequence 21, Appl   |

|    |       |      |        |    |                      |                                    |
|----|-------|------|--------|----|----------------------|------------------------------------|
| 15 | 180.8 | 10.8 | 2557   | 10 | US-09-844-948-1      | Sequence 11, Appl1                 |
| 16 | 180.8 | 10.8 | 2557   | 13 | US-10-168-648-19     | Sequence 19, Appl1                 |
| 17 | 180.8 | 10.8 | 2566   | 10 | US-09-814-353-19076  | Sequence 19076, A                  |
| 18 | 166   | 9.0  | 2408   | 16 | US-10-108-260A-730   | Sequence 730, App                  |
| 19 | 150.8 | 9.0  | 1288   | 15 | US-10-072-977-13     | Sequence 13, Appl                  |
| 20 | 113.4 | 6.8  | 500    | 10 | US-09-814-353-12805  | Sequence 12805, A                  |
| 21 | 107.8 | 6.5  | 477    | 10 | US-09-918-995-2051   | Sequence 2051, Ap                  |
| 22 | 102.4 | 6.1  | 502    | 10 | US-09-814-353-422    | Sequence 22, Appl                  |
| 23 | 102.4 | 6.1  | 502    | 10 | US-09-814-353-4420   | Sequence 4420, Ap                  |
| 24 | 52.6  | 3.2  | 1538   | 17 | US-10-437-963-598    | Sequence 598, Ap                   |
| 25 | 45.6  | 2.7  | 955    | 17 | US-10-437-963-17850  | Sequence 17850, A                  |
| 26 | 45.4  | 2.7  | 954    | 13 | US-10-424-599-123113 | Sequence 123113, Sequence 628, App |
| 27 | 44    | 2.6  | 779    | 9  | US-09-925-300-628    | Sequence 5103, Ap                  |
| 28 | 42.4  | 2.5  | 435    | 10 | US-09-814-353-5103   | Sequence 11395, A                  |
| 29 | 42.4  | 2.5  | 435    | 10 | US-09-814-353-11395  | Sequence 37, Appl                  |
| 30 | 42.2  | 2.5  | 17466  | 13 | US-10-235-192A-37    | Sequence 1, Appl1                  |
| 31 | 42.2  | 2.5  | 174566 | 15 | US-10-020-111-1      | Sequence 2, Appl1                  |
| 32 | 41.6  | 2.4  | 1170   | 13 | US-09-916-841-2      | Sequence 15, Appl                  |
| 33 | 40.6  | 2.4  | 431    | 17 | US-10-437-963-33254  | Sequence 33254, A                  |
| 34 | 40.2  | 2.4  | 431    | 17 | US-10-060-036-1271   | Sequence 2171, Ap                  |
| 35 | 40    | 2.4  | 150    | 15 | US-09-834-975-47     | Sequence 147, Appl                 |
| 36 | 40    | 2.4  | 368    | 9  | US-09-960-352-1817   | Sequence 1817, A                   |
| 37 | 40    | 2.4  | 345    | 10 | US-09-814-353-5466   | Sequence 5466, Ap                  |
| 38 | 39.8  | 2.4  | 345    | 10 | US-09-814-353-11753  | Sequence 11753, A                  |
| 39 | 39.8  | 2.4  | 1775   | 11 | US-09-764-875-170    | Sequence 170, App                  |
| 40 | 39.8  | 2.4  | 1650   | 13 | US-10-425-114-10505  | Sequence 30550, A                  |
| 41 | 39.8  | 2.4  | 1833   | 13 | US-10-425-114-13904  | Sequence 13904, A                  |
| 42 | 39.8  | 2.4  | 234    | 15 | US-09-764-846-19     | Sequence 19, Appl                  |
| 43 | 39.6  | 2.4  | 234    | 9  | US-10-091-443-19     | Sequence 15, Appl                  |
| 44 | 39.6  | 2.4  | 413    | 9  | US-09-880-107-226    | Sequence 226, App                  |
| 45 | 39.6  | 2.4  | 413    | 9  | US-09-880-107-226    | Sequence 236, App                  |

## ALIGNMENTS

```

RESULT 1
US-10-080-960-1
Sequence 1, Application US/10080960
Publication No. US20020197695A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Glucksmann, Maria
APPLICANT: Meyers, Rachel
TITLE OF INVENTION: 80090, 52874, 52880, 63497, AND 33425
TITLE OF INVENTION: METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES THEREOF
FILE REFERENCE: 38155-20044.00
CURRENT APPLICATION NUMBER: US/10/080,960
CURRENT FILING DATE: 2001-10-19
PRIOR APPLICATION NUMBER: US 60/242,040
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: US 60/242,038
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: US 60/241,992
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: US 60/242,637
PRIOR FILING DATE: 2000-10-23
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1669
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (163)...(1623)
US-10-080-960-1
100.0%; Score 1669; DB 14; Length 1669;
Query Match
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 CAGCGCTCGCGCTCTGCTGCTTAGTGTGACTTTGCGCTCTAGGTGATGCATGACTTT 60

```





```

; LOCATION: 1
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20757

Query Match      97.2%; Score 1622; DB 10; Length 2937;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1642; Conservative 0; Mismatches 10; Indels 4; Gaps 1;

QY 1 CACGCGTCCGCTCTGCTCTAGTGTGACTTTGGGCTCAGGTGATCCATGACTTTT 60
DB 2937 CACGCGTCCGCTCTGCTCTAGTGTGACTTTGGGCTCAGGTGATCCATGACTTTT 2878
QY 61 TAAAGCCAAATATATTTCTTACTCTTTGAGAGTGTGCTTGGCTTTTCACTCAGTGT 120
DB 2877 TAAAGCCAAATATATTTCTTACTCTTTGAGAGTGTGCTTGGCTTTTCACTCAGTGT 2818
QY 121 TTTTCTTTTCTTTTCTTTTGGCCTTGATGACGCTTGAGATCTTAATGAAAGTACAGGGCCCT 180
DB 2817 TTTTCTTTTCTTTTCTTTTGGCCTTGATGACGCTTGAGATCTTAATGAAAGTACAGGGCCCT 2758
QY 181 CCCCAAGGAGTTACAGACTCCATGCAATGCTTCAATGATCAGTGGCTTTATCTTAACCC 240
DB 2757 CCCCAAGGAGTTACAGACTCCATGCAATGCTTCAATGATCAGTGGCTTTATCTTAACCC 2698
QY 241 AGGAGCAGGAGACATATAAAGAGTCTATGCTGAGCTGGGGAAGTTTGAAGGAAAGAG 300
DB 2697 AGGAGCAGGAGACATATAAAGAGTCTATGCTGAGCTGGGGAAGTTTGAAGGAAAGAG 2838
QY 301 TTTAAAGTTCAGTTCGACAGATGACATACAAATATGAGAGAGACCTACGACATCTT 360
DB 2637 TTTAAAGTTCAGTTCGACAGATGACATACAAATATGAGAGAGACCTACGACATCTT 2578
QY 361 AATTCATTTCTTAAAGAAAGAGATTGACTTCAACAGGAAAGAAATGGGAAATGGAC 420
DB 2577 AATTCATTTCTTAAAGAAAGAGATTGACTTCAACAGGAGTGA---TGGAGTGGAC 2522
QY 421 AGCTACCCCATATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
DB 2521 AGCTACCCCATATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2462
QY 481 TGTGAGCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 2461 TGTGAGCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2402
QY 541 AAAAGCTTCTCTTCTTAAGTACTGACTTTAAAGTACTGACTTTAAAGTACTGACTTTAA 600
DB 2401 AAAAGCTTCTCTTCTTAAGTACTGACTTTAAAGTACTGACTTTAAAGTACTGACTTTAA 2342
QY 601 GCCCATCATGACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 2341 GCCCATCATGACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2282
QY 661 CATTAACCAAGTATTCCTGTTCACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
DB 2281 CATTAACCAAGTATTCCTGTTCACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2222
QY 721 CCACATAACCAACTTACTGAGAGCATTTGAAGTCTGGAAGTCACTCCGATCTCTAGTT 780
DB 2221 CCACATAACCAACTTACTGAGAGCATTTGAAGTCTGGAAGTCACTCCGATCTCTAGTT 2162
QY 781 CCTTTGACGTCACAAAACAAAGCTTGAAGAAAGCTTGTCTGCTGCTGCTGCTGCTGCTGCT 840
DB 2161 CCTTTGACGTCACAAAACAAAGCTTGAAGAAAGCTTGTCTGCTGCTGCTGCTGCTGCTGCT 2102
QY 841 GACTGTGACCCCAACCAAGAGAGAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 2101 GACTGTGACCCCAACCAAGAGAGAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2042
QY 901 GTGATTTCTTAAGTGAATGTTTACGAAACAAAGACTCTCCCTCAGCAGCTGAAAAATCCA 960
DB 2041 GTGATTTCTTAAGTGAATGTTTACGAAACAAAGACTCTCCCTCAGCAGCTGAAAAATCCA 1982
QY 961 GCCTCTATGAGTCCGATGCTTTTATAGATCATTTGACAGTATAGTTTACTCTAGCT 1020

```

```

DB 1981 GCCTCTATGAGTGCAGATGCTTTTATAGATCATTTGACAGATATAGTTTATCTAGCT 1922
QY 1021 TTGAGATGAGAGTGTGTGATGATGACTACATGACTGAGAGTTCGGAGGCCATGAAATG 1080
DB 1921 TTGAGATGAGAGTGTGTGATGATGACTACATGACTGAGAGTTCGGAGGCCATGAAATG 1862
QY 1081 GGGGTAGTCCCTGTATATTAAGAGATCCCAAGATCAAGACTGCTGCTTCCAGTTAACAA 1140
DB 1861 GGGGTAGTCCCTGTATATTAAGAGATCCCAAGATCAAGACTGCTGCTTCCAGTTAACAA 1802
QY 1141 AGTGTATCTCTGTATAGAAATTTCTACCCAGAGGAAATGGCAAGTTCATCAGACGA 1200
DB 1801 AGTGTATCTCTGTATAGAAATTTCTACCCAGAGGAAATGGCAAGTTCATCAGACGA 1742
QY 1201 CTGATCTGTATGACAGATTTGTATAGAGCCTATGTAAATGAAAGTGAAGGCTGAGATC 1260
DB 1741 CTGATCTGTATGACAGATTTGTATAGAGCCTATGTAAATGAAAGTGAAGGCTGAGATC 1682
QY 1261 TCTAACACAGCACTTCTGACAGCTCTCAGGGAACGGAATGGGGAATGCAAGACGTCAAC 1320
DB 1681 TCTAACACAGCACTTCTGACAGCTCTCAGGGAACGGAATGGGGAATGCAAGACGTCAAC 1622
QY 1321 CAGGACATTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
DB 1621 CAGGACATTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1562
QY 1381 AGGCTTCAGAAAAAGGCTTACACCCAAAAAGATGGAGGACAGAAATACCCACTGAGT 1440
DB 1561 AGGCTTCAGAAAAAGGCTTACACCCAAAAAGATGGAGGACAGAAATACCCACTGAGT 1502
QY 1441 TGCCAGAGCCCAAGAGTGTGCTTCTCACTCCAGCTCCAGCTCCAGCTTGAAGTCTTGG 1500
DB 1501 TGCCAGAGCCCAAGAGTGTGCTTCTCACTCCAGCTCCAGCTCCAGCTTGAAGTCTTGG 1442
QY 1501 CGAGAGATGTGATTTCCAGCTTTGAAACATCCAAAGAAAGAGCCAGGACCTPAAGGTGG 1560
DB 1441 CGAGAGATGTGATTTCCAGCTTTGAAACATCCAAAGAAAGAGCCAGGACCTPAAGGTGG 1382
QY 1561 CTGCTGTATAGAAATCAAAATTTTCACTCAAGAGTTTGGGGCTAGTATCAAGGAC 1620
DB 1381 CTGCTGTATAGAAATCAAAATTTTCACTCAAGAGTTTGGGGCTAGTATCAAGGAC 1322
QY 1621 TGATTTCAAAAATGATCAGATGAAACAGAAAAAA 1656
DB 1321 TGATTTCAAAAATGATCAGATGAAACAGAAAAAGAA 1286

RESULT 3
US-09-731-872-43
; Sequence 43, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bouguerelet, Lydie
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNA5 ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78 US3.REG
; CURRENT APPLICATION NUMBER: US/09/731,872
; PRIOR FILING DATE: 1999-12-07
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 60/187,470
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent.pm
; SEQ ID NO 43
; LENGTH: 1942
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 334..1551

```

NAME/KEY: sig\_peptide  
 LOCATION: 334..426  
 OTHER INFORMATION: Von Heijne matrix  
 OTHER INFORMATION: score 4.0554926521937  
 OTHER INFORMATION: seq TVFLVTLQLDLDF/VE  
 US-09-731-872-43

Query Match 50.98; Score 1517.2; DB 9; Length 1942;  
 Best Local Similarity 99.14; Pred. No. 0;  
 Matches 1536; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

|    |      |   |      |
|----|------|---|------|
| Qy | 118  | TTTTTTTTTTTTTTTTTTTTTTGGCTTGGATACCGTTGAGATCTATGAAAGTCACGGGC | 177  |
| Db | 394  | TTTTGCTGTGTACACTCCAGGCTTGATACCGTTGAGATCTATGAAAGTCACGGGC     | 453  |
| Qy | 178  | CTTCCCGAGGAGTTACAGACTCCATGCAATGCTTCAATGATCAGTGGCTTTATCTAAC  | 237  |
| Db | 454  | CTTCCCGAGGAGTTACAGACTCCATGCAATGCTTCAATGATCAGTGGCTTTATCTAAC  | 513  |
| Qy | 238  | ACCGAGGACCGGAGCATATAAGAGTCAATGCTTGAAGTGGGGAAGTTGAAAGAG      | 297  |
| Db | 514  | ACCGAGGACCGGAGCATATAAGAGTCAATGCTTGAAGTGGGGAAGTTGAAAGAG      | 573  |
| Qy | 298  | GAGTTTAAAGTTTCAAGTTTGAAGTGAATACAAAATGAGGAAAGCACTACGCT       | 357  |
| Db | 574  | GAGTTTAAAGTTTCAAGTTTGAAGTGAATACAAAATGAGGAAAGCACTACGCT       | 633  |
| Qy | 358  | CTTATTTATTTCTTAAAGAAAGAGATTGACCTTCAACAGAAAGAAATGGGAATTG     | 417  |
| Db | 634  | CTTATTTATTTCTTAAAGAAAGAGATTGACCTTCAACAGAAAGAAATGGGAATTG     | 693  |
| Qy | 418  | GACAGTACCCCATTTATGCTGTGTGTCCCGCTGACGGGGAGAGTGGAGTTAGC       | 477  |
| Db | 694  | GACAGTACCCCATTTATGCTGTGTGTCCCGCTGACGGGGAGAGTGGAGTTAGC       | 753  |
| Qy | 478  | CAATGTGAGCAGATGCTGTTTCTTCAACATCAACCGGACCTTCAATCATCATG       | 537  |
| Db | 754  | CAATGTGAGCAGATGCTGTTTCTTCAACATCAACCGGACCTTCAATCATCATG       | 813  |
| Qy | 538  | ACCAAGCATTTCTTCTATGATGATCTGATTAATGATGATGATGATGATGATG        | 597  |
| Db | 814  | ACCAAGCATTTCTTCTATGATGATCTGATTAATGATGATGATGATGATGATG        | 873  |
| Qy | 598  | AAAGCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG        | 657  |
| Db | 874  | AAAGCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG        | 933  |
| Qy | 658  | TTTTCATTAACCATGATTAATGATGATGATGATGATGATGATGATGATGATG        | 717  |
| Db | 934  | TTTTCATTAACCATGATTAATGATGATGATGATGATGATGATGATGATGATG        | 993  |
| Qy | 718  | TTTTCATTAACCATGATTAATGATGATGATGATGATGATGATGATGATGATG        | 777  |
| Db | 994  | TTTTCATTAACCATGATTAATGATGATGATGATGATGATGATGATGATGATG        | 1053 |
| Qy | 778  | GTTTCCTTTGAGTCCAAAAACAAGCTTGAAGAAAGCTTCCGCTGTGTATGTA        | 837  |
| Db | 1054 | GTTTCCTTTGAGTCCAAAAACAAGCTTGAAGAAAGCTTCCGCTGTGTATGTA        | 1113 |
| Qy | 838  | TCAGCTGTGACCCCATGATGATGATGATGATGATGATGATGATGATGATG          | 897  |
| Db | 1114 | TCAGCTGTGACCCCATGATGATGATGATGATGATGATGATGATGATGATG          | 1173 |
| Qy | 898  | GAGTGTGATTTCTATGATGATGATGATGATGATGATGATGATGATGATG           | 957  |
| Db | 1174 | GAGTGTGATTTCTATGATGATGATGATGATGATGATGATGATGATGATG           | 1233 |
| Qy | 958  | CCAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG         | 1017 |
| Db | 1234 | CCAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG         | 1293 |
| Qy | 1018 | GCTTTTGAAGATGCAATTTGTGATGATGATGATGATGATGATGATGATGATG        | 1077 |

|    |      |  |      |
|----|------|--|------|
| Db | 1294 | GCTTTTGAAGATGCAATTTGTGATGATGATGATGATGATGATGATGATGATG | 1353 |
| Qy | 1078 | CTGGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG    | 1137 |
| Db | 1354 | CTGGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG    | 1413 |
| Qy | 1138 | AAAAGTCTATTTCTTGTATGATGATGATGATGATGATGATGATGATGATG   | 1197 |
| Db | 1414 | AAAAGTCTATTTCTTGTATGATGATGATGATGATGATGATGATGATGATG   | 1473 |
| Qy | 1198 | CGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG     | 1257 |
| Db | 1474 | CGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG     | 1532 |
| Qy | 1258 | ATCTTAAACGAGCCTTCTGACAGCTCTCAGGGAACGGAATGGGAGTGAAG   | 1317 |
| Db | 1533 | ATCTTAAACGAGCCTTCTGACAGCTCTCAGGGAACGGAATGGGAGTGAAG   | 1592 |
| Qy | 1318 | AAACGAGCAATTTACATGATGATGATGATGATGATGATGATGATGATG     | 1377 |
| Db | 1593 | AAACGAGCAATTTACATGATGATGATGATGATGATGATGATGATGATG     | 1652 |
| Qy | 1378 | ATCAGGCTTCAAGAAAGGCTTACACCCAAAAGATGGAGGAGCAAGATACCA  | 1437 |
| Db | 1653 | ATCAGGCTTCAAGAAAGGCTTACACCCAAAAGATGGAGGAGCAAGATACCA  | 1712 |
| Qy | 1438 | AGTTCGCCAGAGCCCAAGTGTGCTTCTTCAACCTCCGAGCTCCAGCTTTG   | 1497 |
| Db | 1713 | AGTTCGCCAGAGCCCAAGTGTGCTTCTTCAACCTCCGAGCTCCAGCTTTG   | 1772 |
| Qy | 1498 | TTGCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG     | 1557 |
| Db | 1773 | TTGCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG     | 1832 |
| Qy | 1558 | TGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG      | 1617 |
| Db | 1833 | TGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG      | 1892 |
| Qy | 1618 | GACTGATTTCAAAAATGATGATGATGATGATGATGATGATGATGATG      | 1667 |
| Db | 1893 | GACTGATTTCAAAAATGATGATGATGATGATGATGATGATGATGATG      | 1942 |

RESULT 4  
 US-09-876-997-43  
 ; Sequence 43, Application US/09876997  
 ; Publication No. US20030152921A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Mline Edwards, Jean Baptiste  
 ; APPLICANT: Bouguerelet, Lydie  
 ; APPLICANT: Jobert, Severin  
 ; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS  
 ; FILE REFERENCE: 78 US4, CIP  
 ; CURRENT APPLICATION NUMBER: US/09/876, 997  
 ; CURRENT FILING DATE: 2001-06-08  
 ; PRIOR APPLICATION NUMBER: US 09/731, 872  
 ; PRIOR FILING DATE: 2000-12-07  
 ; PRIOR APPLICATION NUMBER: US 60/187, 470  
 ; PRIOR FILING DATE: 2000-03-06  
 ; PRIOR APPLICATION NUMBER: US 60/169, 629  
 ; PRIOR FILING DATE: 1999-12-08  
 ; NUMBER OF SEQ ID NOS: 482  
 ; SOFTWARE: Patent .pm  
 ; SEQ ID NO 43  
 ; LENGTH: 1942  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 334..1551  
 ; NAME/KEY: sig\_peptide  
 ; LOCATION: 334..426  
 ; OTHER INFORMATION: Von Heijne matrix

|    |      |   |      |
|----|------|---|------|
| Db | 1354 | CTGGGGGGTAGTCCCTGTATATTACCGAATCCCCACGACATCAAGACTGGCTTCCAAAGTAAC | 1413 |
| Qy | 1138 | AAAAGGCAATTCCTGTATACAGATTTCTCAACCCACAGGAACCTGGCAAGTTACATCAGA    | 1197 |
| Db | 1414 | AAAAGGCAATTCCTGTATACAGATTTCTCAACCCACAGGAACCTGGCAAGTTACATCAGA    | 1473 |
| Qy | 1198 | CGACTGATTCGTATGACAGATTGTATGAGGCTTATGAAATGGAACTGAAGGGTGAG        | 1257 |
| Db | 1474 | CGACTGATTCGTATGACAGATTGTATGAGGCTTATGAAATGGAACTGAAGGGT-AG        | 1532 |
| Qy | 1258 | ATATCTAACCAAGCACTTTGACACCTCTACAGGACCGGAANTGGGGATCCAAAGCTC       | 1317 |
| Db | 1533 | ATATCTAACCAAGCACTTTGACACCTCTACAGGACCGGAANTGGGGATCCAAAGCTC       | 1592 |
| Qy | 1318 | AACCAAGACAATTACATCGATCATTTGATGTAATGATGATGACCCAAAGTGGGCTAAT      | 1377 |
| Db | 1593 | AACCAAGACAATTACATCGATCATTTGATGTAATGATGATGACCCAAAGTGGGCTAAT      | 1652 |
| Qy | 1378 | ATCAGGCTTCAGAAAAAGGCTTACACCCAAAGATGGAGGCAAGAAATATCCCACTG        | 1437 |
| Db | 1653 | ATCAGGCTTCAGAAAAAGGCTTACACCCAAAGATGGAGGCAAGAAATATCCCACTG        | 1712 |
| Qy | 1438 | AGTTGGCCAGAGGCCACAGTGTGTTGCTTCTCAACACTCCGACCTTCACCTTTGAGCTCT    | 1497 |
| Db | 1713 | AGTTGGCCAGAGGCCACAGTGTGTTGCTTCTCAACACTCCGACCTTCACCTTTGAGCTCT    | 1772 |
| Qy | 1498 | TTCCGAGAAGATGTGATTTCCAGCTTTGAACAATCCAAAGAAAGGCCACAGCACTAAAG     | 1557 |
| Db | 1773 | TTCCGAGAAGATGTGATTTCCAGCTTTGAACAATCCAAAGAAAGGCCACAGCACTAAAG     | 1832 |
| Qy | 1558 | TGGCTGGTGTATGGAATCAAACTTTTCATCTCAAGGTTTGGGGGCTGATATTCAAG        | 1617 |
| Db | 1833 | TGGCTGGTGTATGGAATCAAACTTTTCATCTCAAGGTTTGGGGGCTGATATTCAAG        | 1892 |
| Qy | 1618 | GACGTATTTCAAAAATGATCAGATGAAACAGAAAAAATTTTTTTTTTTTTTTTTTTT       | 1667 |
| Db | 1893 | GACGTATTTCAAAAATGATCAGATGAAACAGAAAAAATTTTTTTTTTTTTTTTTTTT       | 1942 |

```

1      RESULT 5
2      US-10-080-960-3
3      Sequence 3, Application US/10080960
4      Publication No. US20020197695A1
5      GENERAL INFORMATION:
6      APPLICANT: Millennium Pharmaceuticals, Inc.
7      APPLICANT: GlaxoSmithKline, Maria
8      APPLICANT: Meyers, Rachel
9      TITLE OF INVENTION: 800950, 52874, 52880, 63497, AND 33425
10     TITLE OF INVENTION: METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES THEREOF
11     FILE REFERENCE: 3815-2004.00
12     CURRENT APPLICATION NUMBER: US/10/080,960
13     CURRENT FILING DATE: 2001-10-19
14     PRIOR APPLICATION NUMBER: US 60/242,040
15     PRIOR FILING DATE: 2000-10-20
16     PRIOR APPLICATION NUMBER: US 60/242,038
17     PRIOR FILING DATE: 2000-10-20
18     PRIOR APPLICATION NUMBER: US 60/241,992
19     PRIOR FILING DATE: 2000-10-20
20     PRIOR APPLICATION NUMBER: US 60/242,637
21     PRIOR FILING DATE: 2000-10-23
22     NUMBER OF SEQ ID NOS: 37
23     SOFTWARE: FastSeq for Windows Version 4.0
24     SEQ ID NO 3
25     LENGTH: 1461
26     TYPE: DNA
27     ORGANISM: Homo sapiens
28     US-10-080-960-3

```

```
Query Match      87.5%; Score 1461; DB 14; Length 1461;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1461; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```



QY 924 ACGAACAAGACCTCCCTCAGCAGCTGAAAAATCCAGCTCTATGATGCGCATGCTT 983  
DB 300 ACGAACAAGACCTCCCTCAGCAGCTGAAAAATCCAGCTCTATGATGCGCATGCTT 359  
QY 984 TTAATGATCATGACAGATTAATGTTATTCCTAGCTTTTATGATGACAGTTGTGATGA 1043  
DB 360 TTAATGATCATGACAGATTAATGTTATTCCTAGCTTTTATGATGACAGTTGTGATGA 419  
QY 1044 CTACATCATGGAAGTTCTGGAGGCCCTGAACTGGGGGTAGTCCCTGTATTTAGCG 1103  
DB 420 CTGCAATCATGGAAGTTCTGGAGGCCCTGAACTGGGGGTAGTCCCTGTATTTAGCG 479  
QY 1104 ATCCCCAGCATCATCAGACTGGCTCCCAAGTAAACAAAGTCTATTCTTGATCAGATT 1163  
DB 480 ATCCCCAGCATCATCAGACTGGCTCCCAAGTAAACAAAGTCTATTCTTGATCAGATT 539  
QY 1164 TTCTCAACCCAGGGAAGTTCGCAAGTTACATGACAGACTGGATTCTGATGACAGATTG 1221  
DB 540 TTCTCAACCCAGGGAAGTTCGCAAGTTATATGACAGACTGGATTCTGATGACAGATTG 597

## RESULT 7

US-09-814-353-16082  
/ Sequence 16082, Application US/09814353  
/ Publication No. US20030165831A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Lee, John  
/ APPLICANT: Thompson, Pamela  
/ TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
/ TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
/ FILE REFERENCE: MRI-006B  
/ CURRENT APPLICATION NUMBER: US/09/814,353  
/ PRIOR FILING DATE: 2001-03-21  
/ PRIOR APPLICATION NUMBER: US 60/191,031  
/ PRIOR FILING DATE: 2000-03-21  
/ PRIOR APPLICATION NUMBER: US 60/207,124  
/ PRIOR FILING DATE: 2000-05-25  
/ PRIOR APPLICATION NUMBER: US 60/211,940  
/ PRIOR FILING DATE: 2000-06-15  
/ PRIOR APPLICATION NUMBER: US 60/216,820  
/ PRIOR FILING DATE: 2000-07-07  
/ PRIOR APPLICATION NUMBER: US 60/220,661  
/ PRIOR FILING DATE: 2000-07-25  
/ PRIOR APPLICATION NUMBER: US 60/257,672  
/ PRIOR FILING DATE: 2000-12-21  
/ NUMBER OF SEQ ID NOS: 22037  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 16082  
/ LENGTH: 530  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
US-09-814-353-16082

Query Match 27.1%; Score 453; DB 10; Length 530;  
Best Local Similarity 98.9%; Pred. No. 1e-127;  
Matches 456; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 825 GGTGATGATGACAGTGTGACCTGACCCCATCAGACAGGAGCAGTATGTTGCGGAGCT 884  
DB 70 GGGCGAGGTACAGTGTGACCTGACCCCATCAGACAGGAGCAGTATGTTGCGGAGCT 129  
QY 885 GATGACTTACATGAGGTGATTCCTATGTTGAAATGTTTACGAAACAAAGACCTCCCTCA 944  
DB 130 GATGACTTACATGAGGTGATTCCTATGTTGAAATGTTTACGAAACAAAGACCTCCCTCA 189  
QY 945 GCAGCTGAAAAATCAGGCTGTATGATGCGGAGGCTTTTATGATCATTGACAGTA 1004  
DB 190 GCAGCTGAAAAATCAGGCTGTATGATGCGGAGGCTTTTATGATCATTGACAGTA 249  
QY 1005 TAAGTTTATCTAGCTTTTGAAGATGAGTGTGATGACTACATCACTGAGAAGTTCTG 1064

DB 250 TAAGTTTATCTAGCTTTTGAAGATGAGTGTGATGACTACATCACTGAGAAGTTCTG 309  
QY 1065 GAGGCACTGAAAGTGGGGGTAGTCCCTGTATATTAAGATATCCCGACATCAAGACTG 1124  
DB 310 GAGGCACTGAAAGTGGGGGTAGTCCCTGTATATTAAGATATCCCGACATCAAGACTG 369  
QY 1125 GCTTCCAGTAAACAAAGTGTATTTCTTGTATGAGATTTTTCACCCCGGGAAGCTGGC 1184  
DB 370 GCTTCCAGTAAACAAAGTGTATTTCTTGTATGAGATTTTTCACCCCGGGAAGCTGGC 429  
QY 1185 AAGTTACATCAGACAGCTGATTTCTGATGACAGATTTGTATGAGGCTTATGAAATGAA 1244  
DB 430 AAGTTACATCAGACAGCTGATTTCTGATGACAGATTTGTATGAGGCTTATGAAATGAA 489  
QY 1245 GCTGAAGGTTAGATCTCTTAACCGACGACTTTTGCACGCTC 1285  
DB 490 GCTGAAGGTTAGATCTCTTAACCGACGACTTTTGCACGCTC 530

## RESULT 8

US-09-814-353-3379  
/ Sequence 3379, Application US/09814353  
/ Publication No. US20030165831A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Lee, John  
/ APPLICANT: Thompson, Pamela  
/ TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
/ TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
/ FILE REFERENCE: MRI-006B  
/ CURRENT APPLICATION NUMBER: US/09/814,353  
/ PRIOR FILING DATE: 2001-03-21  
/ PRIOR APPLICATION NUMBER: US 60/191,031  
/ PRIOR FILING DATE: 2000-03-21  
/ PRIOR APPLICATION NUMBER: US 60/207,124  
/ PRIOR FILING DATE: 2000-05-25  
/ PRIOR APPLICATION NUMBER: US 60/211,940  
/ PRIOR FILING DATE: 2000-06-15  
/ PRIOR APPLICATION NUMBER: US 60/216,820  
/ PRIOR FILING DATE: 2000-07-07  
/ PRIOR APPLICATION NUMBER: US 60/220,661  
/ PRIOR FILING DATE: 2000-07-25  
/ PRIOR APPLICATION NUMBER: US 60/257,672  
/ PRIOR FILING DATE: 2000-12-21  
/ NUMBER OF SEQ ID NOS: 22037  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 3379  
/ LENGTH: 495  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: 368..444  
/ OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-3379

Query Match 27.0%; Score 450.6; DB 10; Length 495;  
Best Local Similarity 97.5%; Pred. No. 5.5e-127;  
Matches 467; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
QY 825 GGTGATGATGACAGTGTGACCTGACCCCATCAGACAGGAGCAGTATGTTGCGGAGCT 884  
DB 17 GGGCGAGGTACAGTGTGACCTGACCCCATCAGACAGGAGCAGTATGTTGCGGAGCT 76  
QY 885 GATGACTTACATCAGGCTGATTCCTATGTTGAAATGTTTACGAAACAAAGACCTCCCTCA 944  
DB 77 GATGACTTACATCAGGCTGATTCCTATGTTGAAATGTTTACGAAACAAAGACCTCCCTCA 136  
QY 945 GCAGCTGAAAAATCAGGCTGTATGATGCGGAGTGGCTTTTATGATCATTGACAGTA 1004  
DB 137 GCAGCTGAAAAATCAGGCTGTATGATGCGGAGTGGCTTTTATGATCATTGACAGTA 196

QY 1005 TAAATTATCTAGCTTTTGAAGATGCAATTGTGATGATACATCACTGAGAAATTCTG 1064  
DB 197 TAAATTATCTAGCTTTTGAAGATGCAATTGTGATGATGATACATCACTGAGAAATTCTG 256  
QY 1065 GAGGCCACTGAAACTGGGGGT-AGTCCCTGTATATTAGGATCCCCAGCATGACAGACT 1123  
DB 257 GAGGCCACTGAAACTGGGGGTAGTCCCTGTATATTAGGATCCCCAGCATGACAGACT 316  
QY 1124 GGCCTCCAGTAACAAGATGCTATTCTGTATCAGAAATTTCTCAGCCCGAGGAACTGG 1183  
DB 317 GGCCTCCAGTAACAAGATGCTATTCTGTATCAGAAATTTCTCAGCCCGAGGAACTGG 376  
QY 1184 CAAGTTACATCAGACAGCTGATTTGATGACAGATTTGATGAGGCTTATGTAGAAATGA 1243  
DB 377 CAAGTTACATCAGACAGCTGATTTGATGACAGATTTGATGAGGCTTATGTAGAAATGA 436  
QY 1244 AGCTGAAGGTGAGATCTCTTAACGACGACCTTCTCAGGAGCTTCAGGGAACGGAATGG 1302  
DB 437 AGCTGAAGGTGAGATCTCTTAACGACGACCTTCTCAGGAGCTTCAGGGAACGGAATGG 495

## RESULT 9

US-09-814-353-9698  
; Sequence 9698, Application US/09814353  
; Patent No. US2003016581A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MEI-0068  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9698  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 368, 444  
; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-9698

Query Match 27.0%; Score 450.6; DB 10; Length 495;  
Best Local Similarity 97.5%; Pred. No. 5.5e-127;  
Matches 467; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
QY 825 GGTGTATGTACAGTCACTGACCCACATCAGACAGGACAGCTATGTTCCGAGCT 884  
DB 17 GCGCGAGGTACACTGACGTGTACCCACATCAGACAGGACAGCTATGTTCCGAGCT 76  
QY 885 GATGACTTACATGAGGTGATTTCTATGTGTAATGTTTACGAAACAAAGACTTCCTCA 944  
DB 77 GATGACTTACATGAGGTGATTTCTATGTGTAATGTTTACGAAACAAAGACTTCCTCA 136  
QY 945 GCGAGTGAAGAAATCCAGGCTCTATGATGCCAGTGGCTTTATGATCATCTGACAGTA 1004

DB 137 GCAGTGAAGATCCAGCTCTATGATGCCAGATGCTTTTATAGATCATTCGACAGTA 196  
QY 1005 TAAATTATCTAGCTTTTGAAGATGCAATTGTGATGATGATACATCACTGAGAAATTCTG 1064  
DB 197 TAAATTATCTAGCTTTTGAAGATGCAATTGTGATGATGATGATACATCACTGAGAAATTCTG 256  
QY 1065 GAGGCCACTGAAACTGGGGGT-AGTCCCTGTATATTAGGATCCCCAGCATGACAGACT 1123  
DB 257 GAGGCCACTGAAACTGGGGGTAGTCCCTGTATATTAGGATCCCCAGCATGACAGACT 316  
QY 1124 GGCCTCCAGTAACAAGATGCTATTCTGTATCAGAAATTTCTCAGCCCGAGGAACTGG 1183  
DB 317 GGCCTCCAGTAACAAGATGCTATTCTGTATCAGAAATTTCTCAGCCCGAGGAACTGG 376  
QY 1184 CAAGTTACATCAGACAGCTGATTTGATGACAGATTTGATGAGGCTTATGTAGAAATGA 1243  
DB 377 CAAGTTACATCAGACAGCTGATTTGATGACAGATTTGATGAGGCTTATGTAGAAATGA 436  
QY 1244 AGCTGAAGGTGAGATCTCTTAACGACGACCTTCTCAGGAGCTTCAGGGAACGGAATGG 1302  
DB 437 AGCTGAAGGTGAGATCTCTTAACGACGACCTTCTCAGGAGCTTCAGGGAACGGAATGG 495

## RESULT 10

US-09-764-877-187  
; Sequence 187, Application US/09764877  
; Patent No. US20020147140A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005  
; CURRENT APPLICATION NUMBER: US/09/764,877  
; CURRENT FILING DATE: 2001-02-17  
; PRIOR APPLICATION data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 187  
; LENGTH: 378  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (347)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-877-187

Query Match 19.4%; Score 323.8; DB 9; Length 378;  
Best Local Similarity 96.5%; Pred. No. 3.6e-88;  
Matches 363; Conservative 0; Mismatches 8; Indels 5; Gaps 3;  
QY 694 GCGAGTTCAGCAGGATTCCTCCACTGCTACCTACCTACCTACCTACCTACCTACCTACCT 753  
DB 1 GCGAGTTCAGCAGGATTCCTCCACTGCTACCTACCTACCTACCTACCTACCTACCTACCT 60  
QY 754 GTCTGGAAGTACTCCGATACCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 813  
DB 61 GTCTGGAAGTACTCCGATACCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 120  
QY 814 CTGTGCGCTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 873  
DB 121 CTGTGCGCTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
QY 874 GTTCCGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 933  
DB 181 GTTCCGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
QY 934 GACCTCCCTGAGGCTGAAATAATCCAGGCTCTATGATGATGATGATGATGATGATGATGAT 993  
DB 241 GACCTCCCTGAGGCTGAAATAATCCAGGCTCTATGATGATGATGATGATGATGATGATGAT 300  
QY 994 ATTGCAGATTAAGTTTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1053  
DB 301 ATTGCAGATTAAGTTTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 355

QY 1054 GAGAACTTCTGGAGGC 1069  
|  
DB 356 GGAAGTCTGGAGGC 371

RESULT 11  
US-10-242-515-187

/ Sequence 187, Application US/10242515  
/ Publication No. US2004009488A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Rosen et al.  
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
/ FILE REFERENCE: PC005051  
/ CURRENT APPLICATION NUMBER: US/10/242,515  
/ CURRENT FILING DATE: 2002-09-13  
/ PRIOR APPLICATION NUMBER: 09/764,877  
/ PRIOR FILING DATE: 2001-01-17  
/ PRIOR APPLICATION NUMBER: 60/179,065  
/ PRIOR FILING DATE: 2000-01-31  
/ PRIOR APPLICATION NUMBER: 60/180,628  
/ PRIOR FILING DATE: 2000-02-04  
/ PRIOR APPLICATION NUMBER: 60/214,886  
/ PRIOR FILING DATE: 2000-06-28  
/ PRIOR APPLICATION NUMBER: 60/217,487  
/ PRIOR FILING DATE: 2000-07-11  
/ PRIOR APPLICATION NUMBER: 60/225,758  
/ PRIOR FILING DATE: 2000-08-14  
/ PRIOR APPLICATION NUMBER: 60/220,963  
/ PRIOR FILING DATE: 2000-07-26  
/ PRIOR APPLICATION NUMBER: 60/217,496  
/ PRIOR FILING DATE: 2000-07-11  
/ PRIOR APPLICATION NUMBER: 60/225,447  
/ PRIOR FILING DATE: 2000-08-14  
/ PRIOR APPLICATION NUMBER: 60/218,290  
/ PRIOR FILING DATE: 2000-07-14  
/ Remaining Prior Application data removed - See File Wrapper or PALM.  
/ NUMBER OF SEQ ID NOS: 4031  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 187  
/ LENGTH: 378  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ LOCATION: (347)  
/ OTHER INFORMATION: n equals a,t,g, or c  
US-10-242-515-187

Query Match 19.4%; Score 323.8; DB 16; Length 378;  
Best Local Similarity 96.5%; Pred. No. 3.6e-88;  
Matches 363; Conservative 0; Mismatches 8; Indels 5; Gaps 3;

QY 694 GCCAGGTCAGAGGACATTCACCTTGGCACTAATCCCAATCTTGGAGAGCATTTGAA 753  
|  
DB 1 GCCAGGTCAGAGGACATTCACCTTGGCACTAATCCCAATCTTGGAGAGCATTTGAA 60  
754 GTCCTGAAGTCACTCCGATACCTAGTCTTTGGAGTCCAAAAAAGAGCTTAGAAAAA 813  
61 GTCCTGAAGTCACTCCGATACCTAGTCTTTGGAGTCCAAAAAAGAGCTTAGAAAAA 120  
814 CTTGCTCCGCTGGTATATGTACAGTCAACCTGTACCCACCATATAGACAGGACAGCTAT 873  
121 CTTGCTCCGCTGGTATATGTACAGTCAACCTGTACCCACCATATAGACAGGACAGCTAT 180  
874 GTCCTGAAGTCACTCCGATACCTAGTCTTTGGAGTCCAAAAAAGAGCTTAGAAAAA 933  
181 GTTCGAGAGTATACCTTACATGAGGTGATTCCTATGCTGATTTTACGAAAGAAA 240  
934 GACCTTCCTCAGACAGCTGAAAAATCAGCCTCTATGATGCGGATGCTTTTATAGATC 993  
241 GACCTTCCTCAGACAGCTGAAAAATCAGCCTCTATGATGCGGATGCTTTTATAGATC 300

QY 994 ATTGACAGTATTAAGTTTATTCCTAGCTTTTGAGAAATGACAGTTTGTGATGACTACATCACT 1053  
|  
DB 301 A-TGCACAGTATTAAGTTTATTCCTAGCTTTT--GAGATGACTTTGTATGA-NACATCACT 355  
QY 1054 GAGAACTTCTGGAGGC 1069  
|  
DB 356 GGAAGTCTGGAGGC 371

RESULT 12  
US-10-467-595-50

/ Sequence 50, Application US/10467595  
/ Publication No. US20040166501A1  
/ GENERAL INFORMATION:  
/ APPLICANT: AZIMZAI, Yalda; YUE, Henry;  
/ APPLICANT: DING, Li; NGUYEN, Dammel B.;  
/ APPLICANT: GANDHI, Ameena R.; BURFORD, Neil;  
/ APPLICANT: THANAKAVELU, Kavitha; ELLIOTT, Vicki S.;  
/ APPLICANT: RAMKUMAR, Jayalaxmi; YAO, Montique G.;  
/ APPLICANT: LAU, Preeti G.; TANG, Y. Tom;  
/ APPLICANT: SWARNAKAR, Anita; WARREN, Bridget A.;  
/ APPLICANT: WALIA, Narinder K.; POLICKY, Jennifer L.;  
/ APPLICANT: XU, Yuning; HONCHELL, Cynthia D.;  
/ APPLICANT: AU-YOUNG, Janice K.; BAUGHN, Mariah R.;  
/ APPLICANT: DUGGAN, Brendan M.; LU, Dyung Aina M.;  
/ APPLICANT: GIETZEN, Kimberly J.; JACKSON, Jennifer L.;  
/ APPLICANT: RAUWANN, Bridget E.; LU, Yan;  
/ APPLICANT: KARENT, Stephanie K.; TRAN, Uyen K.;  
/ APPLICANT: RICHARDSON, Thomas W.; EMBELING, Brook M.;  
/ APPLICANT: HAPFLIA April J.A.; BURRILL, John D.;  
/ APPLICANT: MARCUS, Gregory A.; ZINGHER, Kurt A.;  
/ TITLE OF INVENTION: RECEPTORS AND MEMBRANE-ASSOCIATED PROTEINS  
/ FILE REFERENCE: PF-0894 PCT  
/ CURRENT APPLICATION NUMBER: US/10/467,595  
/ CURRENT FILING DATE: 2003-08-06  
/ PRIOR APPLICATION NUMBER: PCT/US02/03868  
/ PRIOR FILING DATE: 2002-02-05  
/ PRIOR APPLICATION NUMBER: US 60/267,201  
/ PRIOR FILING DATE: 2001-02-06  
/ PRIOR APPLICATION NUMBER: US 60/269,580  
/ PRIOR FILING DATE: 2001-02-16  
/ PRIOR APPLICATION NUMBER: US 60/282,679  
/ PRIOR FILING DATE: 2001-04-09  
/ PRIOR APPLICATION NUMBER: US 60/288,295  
/ PRIOR FILING DATE: 2001-05-02  
/ PRIOR APPLICATION NUMBER: US 60/348,687  
/ PRIOR FILING DATE: 2002-01-14  
/ NUMBER OF SEQ ID NOS: 86  
/ SOFTWARE: PERL Program  
/ SEQ ID NO 50  
/ LENGTH: 1550  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ OTHER INFORMATION: Incyte ID No: 1953366CBI  
US-10-467-595-50

Query Match 17.9%; Score 298; DB 17; Length 1550;  
Best Local Similarity 100.0%; Pred. No. 7.6e-80;  
Matches 298; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 AGTCAATGTTGAGCTGGGGAAGTTTGAAGGAAGGAGTTTAAAGTTCCAGCTTGCAAG 322  
|  
DB 185 AGTCAATGTTGAGCTGGGGAAGTTTGAAGGAAGGAGTTTAAAGTTCCAGCTTGCAAG 244  
323 ATGACATACAAAATGAGGAAGACCTCCGATCTTAATTAATTTCTTAAGAAAGAG 382  
245 ATGACATACAAAATGAGGAAGACCTCCGATCTTAATTAATTTCTTAAGAAAGAG 304  
QY 383 GATTGACCTTAAACAGAAAAGAAATGGAATTTGAACAGCTACCCCATTTATGCTGT 442  
|

Db 305 GATTGACCTTCAACAGAAAAGAAATGGGAATTGGACAGCTACCCCATATAGCTTGTG 364  
Qy 443 GGTCCCCGCTGACGGGGGAGACTGGAGGTTAGGCGCATGTGGACAGATGCTGTCTT 502  
Db 365 GGTCCCGCTGAGGGGGAGACTGGAGGTTAGGCGCATGTGGACAGATGCTGTCTT 424  
Qy 503 TCACATCAACCGGACCTACCTCCATCATCAGATGACCAAGACATTCCTCTTATG 560  
Db 425 TCACATCAACCGGACCTACCTCCATCATCAGATGACCAAGACATTCCTCTTATG 482

## RESULT 13

US-09-844-948-3  
; Sequence 3, Application US/09844948  
; Publication No. US20030119161A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel A.  
; APPLICANT: Williamson, Mark  
; TITLE OF INVENTION: 32132, A NOVEL FUCOSYLTRANSFERASE FAMILY  
; FILE REFERENCE: 10448-048001  
; CURRENT APPLICATION NUMBER: US/09/844,948  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: US 60/200,604  
; PRIOR FILING DATE: 2000-04-28  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1479  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-844-948-3

Query Match 10.8%; Score 180.8; DB 10; Length 1479;  
Best Local Similarity 51.8%; Pred. No. 6,6e-44;  
Matches 489; Conservative 0; Mismatches 442; Indels 13; Gaps 3;

Qy 440 GGTGTCTCCCGCTGACGGGGGAGACTGGAGGTTAGGCGCAATGTGGACAAATGCTTGT 499  
Db 252 GCTATTCCTCCCGCTGACGGGGGAGACTGGAGGCGCA-TGAGTGTGCGCGCGCGCTGCG 310  
Qy 500 TCTTCACTCAACACCGGACCTACCTCCATCATCAGATGACCAAGACATTCCTCTTATG 559  
Db 311 TGGCGTCCGGGAACCGCGAGGCGTGGAGGACTCGCGGACCGGCGCGCTCTTCAACG 370  
Qy 560 GTATGACTTTAATAGATAGTACTCTCTGCTCGGAAAGCCCATATACTGGGCTG 619  
Db 371 GCACAGACTTCCCGGCTGGGCGCGCGCTGCGCGCTGCGCGACCAAGCTGGGCGC 430  
Qy 620 TTTTTCATGAAGTCCCGGAAACAAATATATAGCTCTTTCATTAACCATGATTAAT 679  
Db 431 TCTTCCACAGAGAGTGGCGCTCCCACTTCTTCTAGGACGCGCGCGCATTCGCGC 490  
Qy 680 TGTTCATCACTACGCGAGTTGACGAGGACTTCCCACTTCCCACTAATCCCAATACT 739  
Db 491 TCTTCAATCTTACCTCCACTTCACTGCGCACTCGGATTAACCCGCTGCGTGAAGTGC 550  
Qy 740 TGGAGGCAATGAGTCCGAGAGTCACTCCGATACCTGATTCCTTTGAGTCCAAATA 739  
Db 551 TGCCCGGAGACCGCTATCT--GCGCGCGCGGCTGCTCCGCGCAATGAGCGCGGAGT 607  
Qy 800 AGCTTGAAGAAAGACTTGTCTGCTGTGTATGTACAGTCAAGTGTGACCCACCATAG 859  
Db 608 GGGCGCGCGCGCTACGCGCGCTGCTATCTGCTACATCACTGAGAGTGCACGCG 667  
Qy 860 ACAGGAGACGCTATGTTCGCGAGCTGATGACTTAATGAGTGCATTCCTATGTGTAAT 919  
Db 668 ACCGGAGCGCTACGTCGCGAGCTTATGCGCAATCCGCTGAGACTCTTACGGGAAT 727  
Qy 920 GTTTAGAAACAAAGACTTCTCT-----CAGCAGCTGAAAATTCAGCCTCTATG 970  
Db 728 GCTGTGAAGATGGGAGCTGCTTACCGCGCGCTACAGAGACAGCCAGCGCCACACCG 787

Qy 971 ATCCGATGCTTTTATAGATCATTCGACAGTATATAGTTATTCCTAGCTTTTGAATG 1030  
Db 788 AGGATCCAGACCTCTGTGCTTCTTGTCCGCTATATGTTCCACTTGGCCCTGMAAATG 847  
Qy 1031 CAGTTTGTGATGACTACATCACTGAGAACTTCTGAGGCCACTGMAAATCTGGGCTAGTC 1090  
Db 848 CCATCTGTAGAGACTACATGACAAAAAAGTGGCGCTCCCATGACACTGGGGCTGTGC 907  
Qy 1091 CTGTATTTAGGATATCCCGGACATGACAGACTGCGCTTCCAGTAAACAAAGCTATTC 1150  
Db 908 CCGTATCCGCGGCTTCTCTCTGTGAGGACTGATGCCGACATCATCTCCGTATCC 967  
Qy 1151 TTGTATCAGAAATTTTCTACCCGAGGAACCTGCGCACTTCACTGACGACTGATTCG 1210  
Db 968 TGATGATGATTTTGAAGTCTCTCTCAGAGCTGGGAGAGTTATTTGACTTTCTGACAGGA 1027  
Qy 1211 ATGACAGATTTGATGAGGCTATGATGAAATGGAAGCTGAGGATGATCTTACACAG 1270  
Db 1028 ATGATGAGGAGTATGAAATACCTGSCATACAGCAACTGGGGGCAATCACCAAT 1087  
Qy 1271 GACTTCTGACAGCTCTCAGGGAACGAAATGGGAGTGCAGAGAGCTCAACCAAGCAAT 1330  
Db 1088 TTCTTCTGATAGTCTGATGAGATCGGAGAGTGGGAGTGAATCCCTTGTCTGCTAAT 1147  
Qy 1331 ACATGATGACTTTGAGTGTATGCTGTGACACCAAGGTGTGGCT 1374  
Db 1148 ACTTCAGGCTTGAAGTGTTCGCTGTGACTACGAATGCT 1191

## RESULT 14

US-10-184-648-21  
; Sequence 21, Application US/10184648  
; Publication No. US20030224376A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel E.  
; APPLICANT: Williamson, Mark  
; APPLICANT: Tsai, Fong-Ying  
; APPLICANT: Hunter, John J.  
; APPLICANT: Macbeth, Kyle J.  
; APPLICANT: Rudolph-Owen, Laura A.  
; APPLICANT: Leiby, Kevin R.  
; APPLICANT: Kapeller-Libermann, Rosana  
; FILE REFERENCE: 10448-192001  
; CURRENT APPLICATION NUMBER: US/10/184,648  
; PRIOR FILING DATE: 2002-06-27  
; PRIOR APPLICATION NUMBER: US 09/815,028  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: PCT/US01/09358  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: US 60/191,964  
; PRIOR FILING DATE: 2000-03-24  
; PRIOR APPLICATION NUMBER: US 09/801,220  
; PRIOR FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: PCT/US01/07269  
; PRIOR FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 60/187,456  
; PRIOR FILING DATE: 2000-03-07  
; PRIOR APPLICATION NUMBER: US 09/816,714  
; PRIOR FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: PCT/US01/09468  
; PRIOR FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: US 60/191,865  
; PRIOR FILING DATE: 2000-03-24  
; PRIOR APPLICATION NUMBER: US 09/844,948  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: PCT/US01/13805  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,604  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 09/861,164  
; PRIOR FILING DATE: 2001-05-18



```

; PRIOR APPLICATION NUMBER: PCT/US01/16292
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,408
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/883,060
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19138
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/212,079
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/962,678
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: PCT/US01/29963
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/235,044
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 09/973,457
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US 60/238,849
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 10/072,285
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: PCT/US02/03736
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/267,494
; PRIOR FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: US 09/817,910
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: PCT/US01/09633
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/192,092
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 09/842,528
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: PCT/US01/40607
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/199,500
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/882,836
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19543
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/211,730
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/882,872
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19153
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/212,077
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-184-648-21

Query Match      10.8%; Score 180.8; DB 13; Length 1479;
Best Local Similarity 51.8%; Pred. No. 6.6e-44;
Matches 489; Conservative 0; Mismatches 442; Indels 13; Gaps 3;

Db      440 GGTGTCCTCCCGTACGGGGGAGACTGGAGAGTGAAGCCATGTGAGCGATGCTTTGTT 499
        253 GCTATTTCCTCCCACTTCCGGGAGACTGCGAGCGCA-TGAGTGTGCGCGGGCGCGCG 310
Qy      500 TTTTCACTCAACACCGGAGCTACTCCATCATCATGACCAAGCATTCCTCTTCTATG 559
        311 TGGCGTCCCGGAACCGCGAGCGCTGAGAGACTGCGGACCGCGCGCTCTTCTTACG 370
Db      560 GTATGACTTTAACTAATAGCTTACTCTGCTCTGAAAAGCCCATCATGACTGGGCTG 619
        371 GCACGAGACTTCGCGCGCTGCGCGCGCGCGCGCGCTGCGCGCGCTGCGCGCAACAGCTGGGCG 430
```

```

Qy      620 TTTTCACTGAAGAGTCCCCGAAAACATTAATAGCTCTTTCATTAACAGATGATTAAGT 679
        421 TCTTCCACGAGAGAGTGGCCCTCTCAACCACTTGTGAGACCAAGCGCCCGGAGCTCGCC 490
Qy      680 TGTTCATTAACAGTCCAGAGCTTGAAGGATTCGCCACTTGAACCTAATACCAATACT 729
        491 TCTTCACTTAACCTCACTTCACTTCACTGAGTGCACCTGATTAACCGGCTGCTGAGTGGC 550
Db      740 TGAAGAGCATTTAAAGTCTGAAAGTCACTCCGATACCTAGTCTTTTGGACATCCAAAACA 799
        551 TGCCCGGAGCGCTAAGT---GGCGCGCGCGGTGCTCGGCCCATAGGAACGCGCGAGT 607
Qy      800 AGCTTAGAAAAGACTTGTCCCGTGTGTATATGATGATGAGAGTGTGACCCACCAATCAG 859
        608 GCGCGCGCGCGCTGACGCGCGCTGCTATATGACGATCACTGAGCTGACGCGCGCGG 667
Db      668 ACCGGACCGCTAAGTGGGAGGCTCATGGCCACATCCCGGTAGACTCTTAGGGAAAT 727
Qy      920 GTTTCGAAAACAAAGACTCTCTT-----CAGCAGCTGAAAATCCAGCTCTATAG 970
        728 GCTTCAGAAATCGGAGAGTGTGCTTACCGCGGCTACAGACAGACGCGCCACCAACCG 787
Qy      971 ATGCCGATGCGCTTTTATAGATCATTCGACGATATAAGTTATCTTAGTTGAGATG 1030
        788 AGGATCCAGAGCTCTTGAGCTTCTTCTTCCGCTTAAGTTCCACTTGGCCCTGGAATG 847
Db      1031 CAGTTTGTGATGACTACATCACTGAGAAGTTCTGGAGCGCACTGAAAATCGGGGTATGTC 1090
        848 CCATCTGTAAAGCATACATGACAGAAAACGTGGCGCTCCCATGCACTTGGGGGTGTCG 907
Qy      1091 CTGATATTAACGATGCCGCCAGCATCAAGAGTGGCTTCAAGTAAACAAAGTGTATTG 1150
        908 CCGGTACCGCGGTCTTCTCTCTGTGAGGAGCTGATGCCGAACAATCATCTCGTCTATCC 967
Qy      1151 TTGATACGAATTTTTCACCCCGAGGAGAGTGGCAATTCATCAGACGATGATTTCTG 1210
        968 TGATGATGATTTTGAAGTCTCAGAAAGCTGGAGAGTTATTAAGTCTTCTGACACAGA 1027
Db      1211 ATGACAGATTGTAGAGCGCTATGTATAGAAAGCTGAAGGTGAGATTCTTAACGAC 1270
        1028 ATGATGAGGATATATATAAATCCTGGCATCAAGCAACTGGGGGCTATCCAACTCAAT 1087
Qy      1271 GACTTCTGACAGCTCTCAAGGAAAGGAAATGGGAGTGCAGAAGCTCAACAGACAT 1330
        1088 TTCTTCTGATGATGCTGAAGATGGGAGTGGGAGTGAATGATCTTGTCTGCTTAAGT 1147
Db      1331 AATGATGATGATTTGAGTGTATGTTGTGACCAAGGTGTGGGCT 1374
Qy      1148 ACCTCACGCGCTTGCAGTGTTCGTCTGTGACTACGAAGTGGCT 1191

RESULT 15
US-09-844-948-1
; Sequence 1, Application US/09844948
; Publication No. US20030119161A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; APPLICANT: Williamson, Mark
; TITLE OF INVENTION: 32132. A NOVEL FUCOSYLTRANSFERASE FAMILY
; TITLE OF INVENTION: MEMBER AND USES THEREFOR
; FIDE REFERENCE: 10448-045001
; CURRENT APPLICATION NUMBER: US/09/844,948
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: US 60/200,604
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2557
; TYPE: DNA
```

Mon Sep 13 09:12:37 2004

us-10-080-960-1.rnpb

Page 12

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (236)...(1711)  
US-09-844-948-1

Query Match 10.8%; Score 180.8; DB 10; Length 2557;  
Best Local Similarity 51.8%; Pred. No. 9,5e-44;  
Matches 489; Conservative 0; Mismatches 442; Indels 13; Gaps 3;

Search completed: September 13, 2004, 04:59:41  
Job time : 646.019 secs

QY 440 GGTGGTCCCGCTGAGGGGAGACTGGAGGTTAGGCCAATGTGAGCAGATGCTTGT 499  
DB 487 GCTATTCCCGCACTTCCGAGACTCGGAGCGCA-TCGAGTGGCGCGCGCTGACG 545  
QY 500 TCTTCAACATCAACCGAAGCTTACATCATGACGCAAGCAATTCCTTTCTATG 559  
DB 546 TGGCGTCCGGGAACCGCGAGGCTGAGGGA-CTCGGGAACGCGCGCTGCTTTACG 605  
QY 560 GTTACTGACTTTAAGATAGATTACTCTGCTCGGAAAGCCATCATGACTGGCTG 619  
DB 606 GCACAGACTTCCCGCGTGGCGCGCCCGCTGCGCGCTGGCGCACGAGCTGGCGC 665  
QY 620 TTTTCATGAGAGTCCCGAATAATTTATAGCTCTTTCATTAACCATGATTA-CT 679  
DB 666 TCCTTCAAGAGAGTGGCCCTTCAACAACTTCTGTGAGCCAGCGCCGGCATCCGC 725  
QY 680 TGTTCACATCACTGCGAGCTTGAGAGGCAATCCCACTTGCCTAACTAACCAAT-CT 739  
DB 726 TCTTCAATCTTACCTCACCCTTCACTGCGCACTCGAATTACCCGCTGCGTGCAGTGC 785  
QY 740 TGGAGAGCATGAGTCCGAGTCACTCCGATACCTAGTTCCTTTGAGTCAAAACA 799  
DB 786 TGCCTGGAGCCGCTATCT--GCGCGCGCGGTGCTCGCCCATGGAACGCGGAGT 842  
QY 800 AGCTTAGAAAAAGCTTGTCCGCTGGTGTATGTAAGTCAAGTCACTGTGACCAATCAG 859  
DB 843 GGGCGCGCGCGCTACGGCGCGCTGTATCTGAGTCACTGCGAGCTGCGAGCGG 902  
QY 860 ACAAGGACAGCTATGTTCCGAGCTGATGACTTACATGAGTGCATTCTATGTAAT 919  
DB 903 ACCGGGACCGCTACGCGCGAGCTCATGCGCACATCCGCTAGACTCTACGGAAT 962  
QY 920 GTTACGAAACAAGACTCCCT-----CAGCAGTGAATAATCCAGCTCTATGG 970  
DB 963 GCCTGAGAAATCGGAGCTGCTACCGCGGCTACAGACACAGCCACGCGCACCG 1022  
QY 971 ATCCCAATGCTTTTATAGATCATGCAAGTAAAGTTTATCTTACCTTTTGAATG 1030  
DB 1023 AGATCCAGAGCTTGTGGCTTTCTTGTCCGCTATAGTTCACCTTGCCCTGAAAAATG 1082  
QY 1031 CAGTTTGTATGACTACATCATGAGAAAGTTGTGAGGCACTGAAACTGGGGGTAGTCC 1090  
DB 1083 CCATCTGTAAAGCTTCAATGACAGAAAACTGTGGCTGCCATGCACTGGGCGCTGTGC 1142  
QY 1091 CTGTATATTAAGGATCCCGACGATCAGAGCTGGCTTCAAGTAAACAAAGTCTATTTC 1150  
DB 1143 CCGTGTACCGCGGTTCTCCCTCTGTGAGGGAGCTGGATGCCGAACATCATCCGCTATCC 1202  
QY 1151 TTGTATCAGATTTTCTACCCCGAGGA-CTGGCAAGTTAATCAGACGACTGGAATCTG 1210  
DB 1203 TGAATGATGATTTTGAAGTCTCTCAGAAAGCTGGCAGAGTTTATGACTTTCTGGAACAAGA 1262  
QY 1211 ATGACAGATTGTATGAGGCTTATGTAAATGGAAGCTGAAGGTGAGATCTCTAACAGC 1270  
DB 1263 ATGATGAGAGTATATGAATACCTGGCATACAGCAACCTGGGGGATACCAACCAAT 1322  
QY 1271 GACTTCTGACAGCTCTCAGGGAACGAAATGGGAGTGCAGAGCTCAACGAGCAATT 1330  
DB 1323 TTCTTCTGATAGTCTGAAGCATCGGAGTGGGAGTGAATGATCTTTGCTGCTTA-CT 1382  
QY 1331 ACATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1374  
DB 1383 ACCTCAACGCGCTTGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1426

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM nucleic - nucleic search, using sw model

Run on: September 12, 2004, 21:42:25 / Search time 3067.12 Seconds  
(without alignments)  
16249.761 Million cell updates/sec

Title: US-10-080-960-1

Perfect score: 1669  
Sequence: 1 cagcgctcgcctctgctgct.....aaaaaaaaaaaaaaaaaa 1669

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 1493109276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

```
Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description        |
|------------|--------|-------------|--------|-------|--------------------|
| 1          | 1131.4 | 67.8        | 1212   | 29    | AY405135 Homo sapi |
| 2          | 1117.4 | 67.0        | 1212   | 29    | AY405136 Pan trogl |
| 3          | 910.4  | 54.5        | 1041   | 12    | BM559087 AGENCOURT |
| 4          | 800.2  | 47.9        | 1212   | 29    | AY405137 Mus muscu |

| 5  | 794   | 47.6 | 1201 | 9  | AL528778  | AL528778  | AL528778  |
|----|-------|------|------|----|-----------|-----------|-----------|
| 6  | 694.4 | 41.6 | 712  | 12 | BM717062  | BM717062  | UI-B-EJO- |
| 7  | 686   | 41.1 | 960  | 13 | BM348736  | BM348736  | BM348736  |
| 8  | 561   | 33.6 | 687  | 12 | BM332058  | BM332058  | 60243246  |
| 9  | 448.8 | 26.9 | 751  | 14 | CF748965  | CF748965  | UI-M-HUO- |
| 10 | 444   | 25.6 | 945  | 10 | BE718163  | BE718163  | 60180438  |
| 11 | 422.6 | 25.3 | 562  | 10 | BE7181825 | BE7181825 | 204047 MA |
| 12 | 415.2 | 24.9 | 849  | 13 | BU261379  | BU261379  | 603374248 |
| 13 | 390   | 23.4 | 523  | 12 | BM090273  | BM090273  | 505452 MA |
| 14 | 388.8 | 23.3 | 728  | 13 | BU449301  | BU449301  | 603764226 |
| 15 | 350.6 | 21.0 | 839  | 13 | BM873167  | BM873167  | 603764226 |
| 16 | 349.4 | 20.9 | 522  | 9  | AL599664  | AL599664  | DKFZD13C  |
| 17 | 347.8 | 20.8 | 555  | 10 | BE665357  | BE665357  | 154293 MA |
| 18 | 346.2 | 20.7 | 724  | 13 | BU300928  | BU300928  | 60373330  |
| 19 | 328.4 | 19.7 | 871  | 12 | BF78017   | BF78017   | UI-M-HDO- |
| 20 | 328   | 19.7 | 771  | 14 | CF738017  | CF738017  | UI-M-HDO- |
| 21 | 326.4 | 19.6 | 672  | 13 | BU319662  | BU319662  | 603510787 |
| 22 | 323.4 | 19.4 | 820  | 13 | BU440303  | BU440303  | 603019157 |
| 23 | 320.2 | 19.2 | 490  | 9  | AL140437  | AL140437  | qaz7a06.X |
| 24 | 319   | 19.1 | 707  | 13 | BU476024  | BU476024  | 603846742 |
| 25 | 315.2 | 18.9 | 595  | 13 | BU464928  | BU464928  | 603776201 |
| 26 | 307   | 18.4 | 1031 | 13 | BU449854  | BU449854  | 603218748 |
| 27 | 297.2 | 17.8 | 696  | 29 | EX150379  | EX150379  | Danio rer |
| 28 | 296.4 | 17.8 | 431  | 14 | CB546903  | CB546903  | AMGNNUC.N |
| 29 | 287   | 17.2 | 501  | 28 | AO338477  | AO338477  | HS 3118.B |
| 30 | 285.2 | 17.1 | 918  | 13 | BQ725661  | BQ725661  | AGENCOURT |
| 31 | 282.2 | 16.9 | 938  | 13 | BU514180  | BU514180  | AGENCOURT |
| 32 | 275   | 16.5 | 865  | 13 | BQ893273  | BQ893273  | AGENCOURT |
| 33 | 270.2 | 16.2 | 806  | 12 | B1739095  | B1739095  | 603361351 |
| 34 | 269.2 | 16.1 | 985  | 10 | BE745086  | BE745086  | 601576194 |
| 35 | 267.6 | 16.0 | 912  | 13 | BU128147  | BU128147  | 603114758 |
| 36 | 264.4 | 15.8 | 791  | 14 | CB988679  | CB988679  | AGENCOURT |
| 37 | 260.4 | 15.6 | 292  | 9  | AL693529  | AL693529  | w43105.X  |
| 38 | 242.2 | 14.5 | 1044 | 13 | BM349997  | BM349997  | EX349997  |
| 39 | 240.6 | 14.4 | 412  | 12 | BM254238  | BM254238  | 515603 MA |
| 40 | 239.6 | 14.4 | 924  | 12 | B1755527  | B1755527  | 603027388 |
| 41 | 221.6 | 13.3 | 239  | 10 | BF334244  | BF334244  | RC1-CT024 |
| 42 | 216.2 | 13.0 | 699  | 13 | BU610965  | BU610965  | UI-M-FCO- |
| 43 | 212.6 | 12.7 | 812  | 13 | BU80453   | BU80453   | 603861250 |
| 44 | 210.4 | 12.6 | 744  | 13 | BM856607  | BM856607  | EX856607  |
| 45 | 207.2 | 12.4 | 798  | 14 | CF738105  | CF738105  | UI-M-HDO- |

#### ALIGNMENTS

RESULT 1  
AY405135  
LOCUS  
DEFINITION Homo sapiens Funt0 gene, VIRUAL TRANSCRIPT, partial sequence.  
ACCESSION AY405135  
VERSION AY405135.1 GI:39761109  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS  
1 (bases 1 to 1212)  
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,  
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE  
Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL  
Science 302 (5652), 1960-1963 (2003)  
PUBMED  
14671302  
2 (bases 1 to 1212)  
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,  
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE  
Direct Submission

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.

FEATURES

source 1..1212

location/Qualifiers

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

<1..>1212

/gene="FUT10"

/locus\_tag="HGM2123"

ORIGIN

Query Match 67.8%; Score 1131.4; DB 29; Length 1212;

Best Local Similarity 99.9%; Pred.No.3.2e-166;

Matches 1132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 263 AGGTGATGTTGAGTGGGAGAGTTGAAAGAGAGAGTTTAAAGTCCAGTTGCAAG 322

DB 80 AGGTGATGTTGAGTGGGAGAGTTGAAAGAGAGAGTTTAAAGTCCAGTTGCAAG 139

QY 323 ATGACATACAAAAATGAGAGAACACCTACGCACTTATTCTTAAAGAAAG 382

DB 140 ATGACATACAAAAATGAGAGAACACCTACGCACTTATTCTTAAAGAAAG 199

QY 383 GATTGACCTTCAACAGAGAAAGAAATGGGAATTTGACAGACCTACCCCTTAAGCTGGT 442

DB 200 GATTGACCTTCAACAGAGAAAGAAATGGGAATTTGACAGACCTACCCCTTAAGCTGGT 255

QY 443 GGTCCCGCTGACGGGGAGACTGGAGGTTAGGCCAATGTGGAGAGAGTCTGTCT 502

DB 260 GGTCCCGCTGACGGGGAGAGACTGGAGGTTAGGCCAATGTGGAGAGAGTCTGTCT 319

QY 503 TCACCATCAACCGGAGCTTACCTCCATCATCATACATGACCAAGACATCTCTCTATAGTA 562

DB 320 TCACCATCAACCGGAGCTTACCTCCATCATCATACATGACCAAGACATCTCTCTATAGTA 379

QY 563 CTGACTTTAACAATAGAGTTAAGCTCTGCTCGGAAAGCCCATGAGTGGGCTGTT 622

DB 380 CTGACTTTAACAATAGAGTTAAGCTCTGCTCGGAAAGCCCATGAGTGGGCTGTT 439

QY 623 TTGATGAGAGTCCCCGAAAAACAATTATTAAGCTCTTTCATTAACCATGATTACCTTGT 682

DB 440 TTGATGAGAGTCCCCGAAAAACAATTATTAAGCTCTTTCATTAACCATGATTACCTTGT 499

QY 683 TCACATACATGCGCAGCTTCAGAGGAGCTTCCACATGCGACATACCTACCAATACCTGG 742

DB 500 TCACATACATGCGCAGCTTCAGAGGAGCTTCCACATGCGACATACCTACCAATACCTGG 559

QY 743 AGAGCATGGAAGTCTGGAAGTCACTCCGATACCTTAAGTCTTTGCAATCCAAAAAAG 802

DB 560 AGAGCATGGAAGTCTGGAAGTCACTCCGATACCTTAAGTCTTTGCAATCCAAAAAAG 619

QY 803 TTGAAAAAAGACTGTCGCGCTGGTATGTACAGTCAAGAGTCAACCCACACATCAGACA 862

DB 620 TTGAAAAAAGACTGTCGCGCTGGTATGTACAGTCAAGAGTCAACCCACACATCAGACA 679

QY 863 GGGACAGCTATGTTGCGAGCTGATGATCTTAACATGAGGTGCAATTCCTATGATGAT 922

DB 680 GGGACAGCTATGTTGCGAGCTGATGATCTTAACATGAGGTGCAATTCCTATGATGAT 739

QY 923 TAGGAACAAAGACCTCCCTCAGACAGTGAATAATCCAGCCCTTATGATCCGATGGCT 982

DB 740 TAGGAACAAAGACCTCCCTCAGACAGTGAATAATCCAGCCCTTATGATCCGATGGCT 799

QY 983 TTATAGGATCATGAGACAGTATTAAGTTATCTTAAGCTTTTGAATCAGATTGTGATG 1042

DB 800 TTATAGGATCATGAGACAGTATTAAGTTATCTTAAGCTTTTGAATCAGATTGTGATG 859

QY 1043 ACTACATCATGGAAGTTCTGGAAGCCCTGAAGCTGGGGGTAGTCCCTGTATATACG 1102

DB 860 ACTACATCATGGAAGTTCTGGAAGCCCTGAAGCTGGGGGTAGTCCCTGTATATACG 919

QY 1103 GATCCCCCAGCATACAGACTGGCTTCCAGTAAACAAAGTCTATTCTGTACAGAT 1162

DB 920 GATCCCCCAGCATACAGACTGGCTTCCAGTAAACAAAGTCTATTCTGTACAGAT 979

QY 1163 TTCTCACCACAGGAATGCGCAAGTTATACAGACAGCTGATTTGATGACAGATTGT 1222

DB 980 TTCTCACCACAGGAATGCGCAAGTTATACAGACAGCTGATTTGATGACAGATTGT 1039

QY 1223 ATGAGGCTATGTAGATATGAACTGAAGGTTGATCTTACACAGACCTTCTGACAG 1282

DB 1040 ATGAGGCTATGTAGATATGAACTGAAGGTTGATCTTACACAGACCTTCTGACAG 1099

QY 1283 CTCTCAGGGAAACGAAATGGGAGTGCAGACGTCACACAGACATTCATCATGATGCA 1342

DB 1100 CTCTCAGGGAAACGAAATGGGAGTGCAGACGTCACACAGACATTCATCATGATGCA 1159

QY 1343 TTGAGTGTATGTGTGCACCAAGGTGTGGGCTAATATACGCTTCAGGAAAG 1395

DB 1160 TTGAGTGTATGTGTGCACCAAGGTGTGGGCTAATATACGCTTCAGGAAAG 1212

RESULT 2

AY405136 1212 bp DNA linear GSS 16-DEC-2003

LOCUS

DEFINITION Pan troglodytes FUT10 gene, VIRUTAL TRANSCRIPT, partial sequence,

ACCESSION

AY405136

VERSION

AY405136.1 GI:39761110

KEYWORDS

GSS.

SOURCE

Pan troglodytes (chimpanzee)

ORGANISM

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE

1 (bases 1 to 1212)

Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sminsky,J.J., Adams,M.D. and Cargill,M.

inferring nonneutral evolution from human-chimp-mouse orthologous gene trios

Science 302 (5652), 1960-1963 (2003)

JOURNAL

Science 302 (5652), 1960-1963 (2003)

PubMed

14671302

2 (bases 1 to 1212)

Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sminsky,J.J., Adams,M.D. and Cargill,M.

Direct Submission

JOURNAL

Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.

FEATURES

source 1..1212

location/Qualifiers

/organism="Pan troglodytes"

/mol\_type="genomic DNA"

/db\_xref="taxon:9598"

<1..>1212

/gene="FUT10"

/locus\_tag="HGM2123"

ORIGIN

Query Match 67.0%; Score 1117.4; DB 29; Length 1212;

Best Local Similarity 98.9%; Pred.No.5e-166;

Matches 1121; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 263 AGGTGATGTTGAGTGGGAGAGTTGAAAGAGAGAGTTTAAAGTCCAGTTGCAAG 322

DB 80 AGGTGATGTTGAGTGGGAGAGTTGAAAGAGAGAGTTTAAAGTCCAGTTGCAAG 139

QY 323 ATGACATACAAAAATGAGAGAACACCTACGCACTTATTCTTAAAGAAAG 382

|    |      |  |      |
|----|------|--|------|
| Dp | 140  | ATGGAATACAAAAATGAGGAGAGCACTACAGCATCTTTAACTTTATCTTTAAGAAAAG     | 199  |
| Qy | 383  | GATTGACCTTCAACAGGAAAAAGAAATGGGAATTGACAGCTTACCCCATTAATGCTTGCT   | 442  |
| Dp | 200  | GATTGACCTTCAACAGGAAAAAGAAATGGGAATTGACAGCTTACCCCATTAATGCTTGCT   | 259  |
| Qy | 443  | GGTCCCCGCTGACGGGGGAGACTGGGAGGTTAAGGCGCAATGAGAGCAAGTGCCTGTTTCT  | 502  |
| Dp | 260  | GGTCCCCGCTGACGGGGGAGACTGGGAGGTTAAGGCGCAATGAGAGCAAGTGCCTGTTTCT  | 319  |
| Qy | 503  | TCACATCAACCGGACCTTACCTCAATCAATGACCAAGGATTCCTTTCTATGTA          | 562  |
| Dp | 320  | TCACATCAACCGGACCTTACCTCAATCAATGACCAAGGATTCCTTTCTATGTA          | 379  |
| Qy | 563  | CTGACCTTTAACTAATATAGCTTAACTCTGCGCTGGAAAGCCCATGATGAGCTGGCTTT    | 622  |
| Dp | 380  | CTGACCTTTAACTAATATAGCTTAACTCTGCGCTGGAAAGCCCATGATGAGCTGGCTTT    | 439  |
| Qy | 623  | TTTCATGAAGTCCCCGAAAAACATTTAAGCTCTTCAATTAACAGATTAATACCTGT       | 682  |
| Dp | 440  | TTTCATGAAGTCCCCGAAAAACATTTAAGCTCTTCAATTAACAGATTAATACCTGT       | 499  |
| Qy | 683  | TCACATCACTGCGCACGTTACAGCAGGCAATTCGCCATTGCGCATTAATCCCAATCTTGG   | 742  |
| Dp | 500  | TCACATCACTGCGCACGTTACAGCAGGCAATTCGCCATTGCGCATTAATCCCAATCTTGG   | 559  |
| Qy | 743  | AGAGCATTTAAGCTCTGAAGTCACTCCGATACCTAGTTCCTTTCAGTCCAAAACAAAC     | 802  |
| Dp | 560  | AGAGCATTTAAGCTCTGAAGTCACTCCGATACCTAGTTCCTTTCAGTCCAAAACAAAC     | 619  |
| Qy | 803  | TTTGAAGAAAGACTTGCTGCTGCGCTGTATGTATGATCAGTCAACTGTGACCCACATCAGAC | 862  |
| Dp | 620  | TTTGAAGAAAGACTTGCTGCTGCGCTGTATGTATGATCAGTCAACTGTGACCCACATCAGAC | 679  |
| Qy | 863  | GGGACAGCTATGTTGGCAGCTGATGATTTCATCGAGGTCGATTCCTTAATGGGAATGT     | 922  |
| Dp | 680  | GGGACAGCTATGTTGGCAGCTGATGATTTCATCGAGGTCGATTCCTTAATGGGAATGT     | 739  |
| Qy | 923  | TACGAAACAAAGACCTCCCTCAGCAGCTGAAAAATCCAGCCTCATGATGCCGATGCT      | 982  |
| Dp | 740  | TACGAAACAAAGACCTCCCTCAGCAGCTGAAAAATCCAGCCTCATGATGCCGATGCT      | 799  |
| Qy | 983  | TTTATATGATCATTTGACAGATTAATGTTATATCCATGCTTTTGAATAGCATTTGTGATG   | 1044 |
| Dp | 800  | TTTATATGATCATTTGACAGATTAATGTTATATCCATGCTTTTGAATAGCATTTGTGATG   | 859  |
| Qy | 1043 | ACTACATCACTGAGAAGTCTGGAAGCCACTGAAACTGCGGGGATAGTCCCTGTATTATAG   | 1102 |
| Dp | 860  | ACTACATCACTGAGAAGTCTGGAAGCCACTGAAACTGCGGGGATAGTCCCTGTATTATAG   | 919  |
| Qy | 1103 | GATCCCCAGCATCACAGCTGGCTTCCAAATTAACAAAGTGCTATTCTTTGATCAAGAT     | 1166 |
| Dp | 920  | GATCCCCAGCATCACAGCTGGCTTCCAAATTAACAAAGTGCTATTCTTTGATCAAGAT     | 979  |
| Qy | 1163 | TTTCTCACCCCGAGGGAATCGCAATTTACATCAGACGACTGGATTCGATGACAGATTGT    | 1222 |
| Dp | 980  | TTTCTCACCCCGAGGGAATCGCAATTTACATCAGACGACTGGATTCGATGACAGATTGT    | 1033 |
| Qy | 1223 | ATGAGGCGCTATGTAATGTAAGCTGAAGGGGTGAGATCTTAAACAGCAGCTTCTGACG     | 1283 |
| Dp | 1040 | ATGAGGCGCTATGTAATGTAAGCTGAAGGGGTGAGATCTTAAACAGCAGCTTCTGACG     | 1099 |
| Qy | 1283 | CTCTCAGGGAACGGAAATGGGGAGTGCAGACGTCACACGAGCAATTAATATGATGATCAT   | 1344 |
| Dp | 1100 | CTCTCAGGGAACGGAAATGGGGAGTGCAGACGTCACACGAGCAATTAATATGATGATCAT   | 1155 |
| Qy | 1343 | TTGAGTGATATGATGTGACCAAGAGTGTGGGCTTAATATATAGGCTTCAGGAAAG        | 1395 |
| Dp | 1160 | TTGAGTGATATGATGTGACCAAGAGTGTGGGCTTAATATATAGGCTTCAGGAAAG        | 1212 |

|                            |  |         |      |        |                 |
|----------------------------|--|---------|------|--------|-----------------|
| LOCUS                      | BM559087   | 1041 bp | mRNA | linear | EST 20-FEB-2000 |
| DEFINITION                 | AGENCOURT.6562690 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5551224   |         |      |        |                 |
| ACCESSION                  | 5', mRNA sequence.   |         |      |        |                 |
| VERSION                    | BM559087   |         |      |        |                 |
| KEYWORDS                   | EST.   |         |      |        |                 |
| SOURCE                     | BM559087.1 GI:18802381   |         |      |        |                 |
| ORGANISM                   | Homo sapiens (human)   |         |      |        |                 |
|                            | Homo sapiens   |         |      |        |                 |
|                            | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  |         |      |        |                 |
|                            | Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.   |         |      |        |                 |
| REFERENCE                  | NIH-MGC http://mgi.nci.nih.gov/.   |         |      |        |                 |
| AUTHORS                    | 1 (bases 1 to 1041)  |         |      |        |                 |
| TITLE                      | National Institutes of Health, Mammalian Gene Collection (MGC)   |         |      |        |                 |
| JOURNAL                    | Unpublished (1999)   |         |      |        |                 |
| COMMENT                    | Contact: Robert Strausberg, Ph.D.<br>Email: gcephs@remail.nih.gov<br>Tissue Procurement: ATCC<br>cDNA Library Preparation: Life Technologies, Inc.<br>cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)<br>DNA Sequencing by: Agencourt Bioscience Corporation<br>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:<br><a href="http://image.lnl.gov">http://image.lnl.gov</a><br>Plate: LHAM12264 row: k column: 01<br>High quality sequence stop: 645.<br>Location/Qualifiers<br>1..1041<br>/organism="Homo sapiens"<br>/mol_type="mRNA"<br>/db_xref="taxon:9606"<br>/clone="IMAGE:5551224"<br>/tissue_type="retinoblastoma"<br>/lab_host="DH10B (phage-resistant)"<br>/clone_id="NIH MGC 67"<br>/note="Organ: eye; Vector: pCMV-Sport6; Site:1: NotI;<br>Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.<br>Average insert size 1.75 kb. Library constructed by Life technologies." |         |      |        |                 |
| ORIGIN                     |  |         |      |        |                 |
| Query Match                | 54.5%; Score 910.4; DB 12; Length 1041;  |         |      |        |                 |
| Best Local Similarity      | 96.5%; Pred. No. 1.ee-133;   |         |      |        |                 |
| Matches 1005; Conservative | 0; Mismatches 27; Indels 9; Gaps 7;  |         |      |        |                 |
| OY                         | 492 TGGTTGTTCTTCACCATCAACCGGACTTACTTCATCATCATATGCCAAGCAATTCT   | 551     |      |        |                 |
| Db                         | 1 TGCTGTTCCTTCACCATCAACCGGACTTACTTCATCATCATATGCCAAGCAATTCT   | 60      |      |        |                 |
| OY                         | 552 CTCTATGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT  | 611     |      |        |                 |
| Db                         | 61 CTCTATGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT   | 120     |      |        |                 |
| OY                         | 612 CTGGGCTGTTTTCATGAAGAGTCCCGAAAAACAATTATTAAGCTCTTCATPAACAGT  | 671     |      |        |                 |
| Db                         | 121 CTGGGCTGTTTTCATGAAGAGTCCCGAAAAACAATTATTAAGCTCTTCATPAACAGT  | 180     |      |        |                 |
| OY                         | 672 GATTACTTGTCACATCACTGACGACGATCAGAGGACATCCCACTTGGCACTAACTAC  | 731     |      |        |                 |
| Db                         | 181 GATCACCTTGTCACATCACTGACGACGATCAGAGGACATCCCACTTGGCACTAACTAC   | 240     |      |        |                 |
| OY                         | 732 CCAATACCTTGAGAGACATTTGAAGTCTCGAAGTCACTCCGATACCTTAGTCTTGGAGTC   | 791     |      |        |                 |
| Db                         | 241 CCAATACCTTGAGAGACATTTGAAGTCTCGAAGTCACTCCGATACCTTAGTCTTGGAGTC   | 300     |      |        |                 |
| OY                         | 792 CAATAAACAGCTTGAATAAAGCTTCTCCGCTGGTGTATGTACAGTCAAGTCAAGTCAAGT   | 851     |      |        |                 |
| Db                         | 301 CAATAAACAGCTTGAATAAAGCTTCTCCGCTGGTGTATGTACAGTCAAGTCAAGTCAAGT   | 360     |      |        |                 |
| OY                         | 852 ACCATGAGACAGGACAGTATGTTCCGAGAGCTATGATCACTATGAGAGTGGATTTCTTA  | 911     |      |        |                 |
| Db                         | 361 ACCATGAGACAGGACAGTATGTTCCGAGAGCTATGATCACTATGAGAGTGGATTTCTTA  | 420     |      |        |                 |
| OY                         | 912 TGGTGAATGTTTACGAAAACAAAGACTCCCTCAGCAGCTGAAAATTCAGACTCTTATGGA   | 971     |      |        |                 |

Db 421 TGGGAATGTTTACGAAACAAAGACCTCCCTCAGCAGCTGAAAGAAATCCAGCCTCTATGGA 480  
Qy 972 TGCAGATGCTTTATAGCATTCATTCAGACATATAAGTTTATCTAGCTTTTGAGAAATGC 1031  
Db 481 TGCGATGGCTTTTATAGCATTCATTCAGACATATAAGTTTATCTAGCTTTTGAGAAATGC 540  
Qy 1032 AGTTTGTGATGACTACATCACTGAGAAATTTCTGAGAGCCACTGAAACTGGGGATGTC 1091  
Db 541 AGTTTGTGATGACTACATCACTGAGAAATTTCTGAGAGCCACTGAAACTGGGGATGTC 600  
Qy 1092 TGTATATTAAGATCCCGCCAGCATCAGACATGAGCTCCAAATGAAAGAGTCTATCT 1151  
Db 601 TGTATATTAAGATCCCGCCAGCATCAGACATGAGCTCCAAATGAAAGAGTCTATCT 660  
Qy 1152 TGTATCAGAAATTTTCTCACCAGGAACTGGCAAGTTACATCAGACATGAGTTCTGA 1211  
Db 661 TGTATCAGAAATTTTCTCACCAGGAACTGGCAAGTTACATCAGACATGAGTTCTGA 720  
Qy 1212 TGAAGATGTAAGAGGCTTATGATGAAATGAAAGCTGAAAGGCTGATCTCTAATCCAGCG 1271  
Db 721 TGAAGATGTAAGAGGCTTATGATGAAATGAAAGCTGAAAGGCTGATCTCTAATCCAGCG 780  
Qy 1272 ACTTCTGACAGCTCTCAGAGGAACGGAATGGGAGTGCAGAGCTC-AACGAGACAAAT 1330  
Db 781 ACTTCTGACAGCTCTCAGAGGAACGGAATGGGAGTGCAGAGCTC-AACGAGACAAAT 840  
Qy 1331 ACATCCATGCA-TTTGATGATGATGCTG-CACCAAGTGTGGCTATATTC-AGGCTTC 1387  
Db 841 ACATCCATGCA-TTTGATGATGATGCTG-CACCAAGTGTGGCTATATTC-AGGCTTC 900  
Qy 1388 AGAAAAAGGCTTACACCCAAAGATGGAGAGCAAGATACCACTGAGTGGCCAG 1447  
Db 901 AGAAAAAGGCTTACACCCAAAGATGGAGAGCAAGATACCACTGAGTGGCCAG 960  
Qy 1448 AGCCCAAGTGTTCG-TTCTCAGCACTCC-GAAGTCCACCTTT--GAGCTCTTTCG 1502  
Db 961 AGCCCAAGTGTTCG-TTCTCAGCACTCC-GAAGTCCACCTTT--GAGCTCTTTCG 1020  
Qy 1503 AGAGATGTGATTTCCAGCTT 1523  
Db 1021 AAAAATTTGAATTTCCAGCTT 1041

RESULT 4  
AY405137 1212 bp DNA linear GSS 16-DEC-2003  
LOCUS Mus musculus FMT10 gene, VIRIDAL TRANSCRIPT, partial sequence,  
DEFINITION Genomic survey sequence.  
ACCESSION AY405137  
VERSION AY405137.1 GI:39761111  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.  
REFERENCE 1 (bases 1 to 1212)  
AUTHORS Clair,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,  
Perriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 1212)  
AUTHORS Clair,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,  
Perriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA

COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source location/Qualifiers  
1..1212  
/organism="Mus musculus"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:10090"  
<1..>1212  
/gene="FMT10"  
/locus\_tag="HCM2123"  
ORIGIN  
Query Match 47.9% Score 800.2; DB 29; Length 1212;  
Best Local Similarity 81.6%; Pred. No. 2.6e-116; Indels 0; Gaps 0;  
Matches 925; Conservative 0; Mismatches 208;  
Qy 263 AGTCATGATGAGCTGGGGAGATTGAAAGAGAGATTAAAGTTCCAGTTTGCAAG 322  
Db 80 AGTAGTGATGAGCTTGCGAATTTGAAAGAGAGATTAAAGACTCCATGTGCAAG 139  
Qy 323 ATGACATACAAAAATGAGAGACACCTACGCTTATTAATGATTTCTTAAGAAAG 382  
Db 140 ACGGCAATGAGATGTGAGGAGAGAGCTTAAACCTGGAGCCATTTCTGAAAAGAG 199  
Qy 383 GATTGACCTTCAACAGAAAAAATGGCAATTGACAGCTACCCCATTAATGCTGTGT 442  
Db 200 CACTGGCTTTGGCTGGAGAGACCAAGGTGATGCTGTACTACCCCATGCTGTGTGT 259  
Qy 443 GGTCCCGCTGACGGGGAGAGCTGGAGATTAGGCCAATGTGAGCAGATGCTTTGTTCT 502  
Db 260 GGTCCCGCTGACGGGGAGAGCTGGAGATTAGGCCAATGTGAGCAGATGCTTTGTTCT 319  
Qy 503 TCACCATCAACGGAGCTACCTCATCATCATGACCAAGATTCCTTCTATAGTA 562  
Db 320 TTACCATCAACGGAGCTTCAGATCATTCGATGACCAAGGCTTCTATTTACAGTA 379  
Qy 563 CTGACTTTAAGATAGATAGCTTACCTGCTCGGTGGAAAGCCCATCATGACTGGGCTGTT 622  
Db 380 CTGACTTTAAGATAGACAGCTTACCTTGTGCTCGGGAAGCCCATCATGACTGGGCTGTT 439  
Qy 623 TTCATGAAGATCCCGGAAAAACATTAATAGCTCTTCAATTAACCGATTAATAGCTGT 682  
Db 440 TCATGAAGAGTCCCGGAAAAACATTAATAGCTCTTCAATTAACCGATTAATAGCTGT 439  
Qy 683 TCACATGACCTGACAGCTTTCAGCAGGATTCACCTTCCACCTACCTACCTACCTTGG 742  
Db 500 TCACATGACCTGACAGCTTTCAGCAGGATTCACCTTCCACCTACCTACCTACCTTGG 559  
Qy 743 AGAGCATGAAGTCTCTGAAGTCACTCCGATACCTAGTCTCTTTCGATGCCAAAAAGC 802  
Db 560 AAGGCGTGAAGTCTCTGAAGTCACTCCGATACCTAGTCTCTTTCGATGCCAAAAAGC 619  
Qy 803 TTAGAAAAAGACTTGTCTCCGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 862  
Db 620 TTAGAAAAAGACTTGTCTCCGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 679  
Qy 863 GGGACAGTATGTTGGCAGAGTGAATGACTTACATGAGGTGATTTCTTATGTTGAATGTT 922  
Db 680 GGGACAGTATGTTGGCAGAGTGAATGACTTACATGAGGTGATTTCTTATGTTGAATGTT 739  
Qy 923 TACGAAACAAAGACCTCCCTCAGAGCTGAAAAATCCAGGCTCTATGATGCGATGGCT 982  
Db 740 TACGAAACAAAGACCTCCCTCAGAGCTGAAAAATCCAGGCTCTATGATGCGATGGCT 799  
Qy 983 TTTATAGGATCATTCAGACAGTATAAGTTTATTCCTAGCTTTTGAATGACATTTGTGATG 1042  
Db 800 TCTACAGAGTCAATGCCAGATATAAGTTTATTCCTAGCTTTTGAATGACATTTGTGATG 859  
Qy 1043 ACTATATCATGAGAAAGTCTGGAGGCACTGAAACCTGGGGGATAGTCCGTATATTAAG 1102  
Db 860 ATTACATCAAGAAAGTCTGGAGGCACTGAAACCTGGGGGATAGTCCGTATATTAAG 919  
Qy 1103 GATCCCGCAGCATCAGAGCTGCTTCCAAAGTAAACAAAGTCTATTTGTATCAGAAAT 1162

```
Db 920 GATCCCCCAACATCCGCTGCTCCAGTAAACCAAGCCGATCTGTGATCGGAAT 979
Qy 1163 TTCTCACCACCAAGGAGTGGCAATTACATCAGACGATGATTTCTGATGACAGATTGT 1222
Db 980 TCTTCACCCCAAGAGAGCTGGCGAACTTCATTAGAGACTGATGATATGACGATGGTTGT 1039
Qy 1223 ATGAGGCTATGTATGATGAGAGCTGAAAGGCTGAGATCTCTAACCCAGCCTTCTGACG 1282
Db 1040 ATGAACTATGTATGATGAGAGCTGAAAGGCGAAGATCTCTAACACGCGCTTCTGACG 1099
Qy 1283 CTCTCAGGGAACGGAATAGGGAGTGCAGAGCTCAACCAAGCAATTATGATGATGAT 1342
Db 1100 CTCTCAGGAAGCGGAGTGGGGTGTGCAGGACATTAAACAGGAACTATGATGACTCGT 1159
Qy 1343 TTGAGTGTATGTGTGTCACCAAGTGTGGGCTATATCAGGCTTCAAGAAAG 1395
Db 1160 TTGAGTGTATGTGTGTGTCGCGAGGGTGTGGCAACAGATGAGCTGACAGAGCAG 1212

RESULT 5
AL528778 1201 bp mRNA linear EST 23-MAY-2003
LOCUS AL528778 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
DEFINITION cDNA clone CSDD001YG24 5'-PRIME, mRNA sequence.
ACCESSION AL528778
VERSION AL528778.2 GI:31066628
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 13, 2001 this sequence version replaced gi:12792271.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqef@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4610.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSDD001BD120P1&cluster=4610.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSDD001BD120P1.
FEATURES
source
location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSDD001YG24"
/issue_type="NEUROBLASTOMA COT 50-NORMALIZED"
/clone_id="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match 47.6%; Score 794; DB 9; Length 1201;
Best Local Similarity 96.8%; Pred. No. 2.5e-115; Indels 2; Gaps 1;
Matches 813; Conservative 7; Mismatches 18;
Qy 259 AAGAGGCTATGTGAGCTGGGAGATTGAAAGGAGAGGTTAAAGTTCCAGTTG 318
Db 307 AACCGGTATGTGAGCTGGGAGATTGAAAGGAGAGGTTAAAGTTCCAGTTG 366
Qy 319 CAAGATGACATACAAATGAGAGACACTGACATCTTAATGATTTCTTAAGAAA 378
```

```
Db 367 CAAGATGACATACAAATGAGAGACACTGACATCTTAATGATTTCTTAAGAAA 426
Qy 379 GAAGATTTGACCTTCAACAGAAAAGAAAATGGAAATTTGACGCTACCCATTANGCTC 438
Db 427 GAAGATTTGACCTTCAACAGAAAAGAAAATGGAAATTTGACGCTACCCATTANGCTC 486
Qy 439 TGGTGTGTCCTCCGCTGAGCGGGGAGACTGGAGGTTAGGCCAATGTGGAGAGATGCTTGT 498
Db 487 TGGTGTGTCCTCCGCTGAGCGGGGAGACTGGAGGTTAGGCCAATGTGGAGAGATGCTTGT 546
Qy 499 TTCTTCAACATCAACCGGACCTTACCTCATCATCAATGACCAAGCAATTCCTTTCTAT 558
Db 547 TTCTTCAACATCAACCGGACCTTACCTCATCATCAATGACCAAGCAATTCCTTTCTAT 606
Qy 559 GGTACTGACTTTAATAGATAGCTTACCTCTGCTCGGAAAGCCATCATGATGAGGGCT 618
Db 607 GGTACTGACTTTAATAGATAGCTTACCTCTGCTCGGAAAGCCATCATGATGAGGGCT 666
Qy 619 GTTTTTCATGAGAGTCCCGGAAAAACAATTATAGCTCTTTTATTAACCAAGATTAAC 678
Db 667 GTTTTTCATGAGAGTCCCGGAAAAACAATTATAGCTCTTTTATTAACCAAGATTAAC 726
Qy 679 TTGTTCACTACACTGCGACGTTTCAGAGGCAATTTCCACTTGGCACTTAATACCANTAC 738
Db 727 TTGTTCACTACACTGCGACGTTTCAGAGGCAATTTCCACTTGGCACTTAATACCANTAC 786
Qy 729 TTGAGAGCATTTGAGTCTGAACTCACTCCGATPACTAGTTCTTTGCACTGCAAAAAC 798
Db 787 TTGAGAGCATTTGAGTCTGAACTCACTCCGATPACTAGTTCTTTGCACTGCAAAAAC 846
Qy 799 AAGCTTAAAAAAGACTTGTCTCGCTGGTGTATGTATGATGACATGTCACCAATCA 858
Db 847 AAGCTTAAAAAAGACTTGTCTCGCTGGTGTATGTATGATGACATGTCACCAATCA 906
Qy 859 GACAGGGAACAGCTATGTTGGAGAGCTGATGATGATGATGATGATGATGATGATGAT 918
Db 907 GACAGGGAACAGCTATGTTGGAGAGCTGATGATGATGATGATGATGATGATGATGAT 966
Qy 919 TGTTTACGAAACAAAGACCTTCCTCAGACAGCTGAAAAATCCAGCTCTATGATGCCGAT 978
Db 967 TGTTTACGAAACAAAGACCTTCCTCAGACAGCTGAAAAATCCAGCTCTATGATGCCGAT 1026
Qy 979 GGCCTTTTATGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1038
Db 1027 GGCCTTTTATGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1086
Qy 1039 GATGACTCATCATCTGAGAGGTTCTGAGGCACTGAAACCTGGGGGATGCTCTGTATAT 1098
Db 1087 GATGACTCATCATCTGAGAGGTTCTGAGGCACTGAAACCTGGGGGATGCTCTGTATAT 1144

RESULT 6
BM717062 712 bp mRNA linear EST 28-FEB-2002
LOCUS BM717062
DEFINITION UI-E-EUO-shk-e-07-0-UI.r2 UI-E-EUO Homo sapiens cDNA clone
UI-E-EUO-shk-e-07-0-UI 5', mRNA sequence.
ACCESSION BM717062.1 GI:19030320
VERSION BM717062.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 712)
Bonaldi, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL MEDLINE
AUTHORS 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
```

375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@iowa.edu

Tissue Procurement: Dr. Gregg Hageman  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
Seq primer: M13 Reverse

## FEATURES

Source

1..712  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-E10-abk-e-07-0-UI"  
/tissue\_type="fetal eyes; lens, eye anterior segment, optic nerve, retina, Retina foveal and Macular, RPE and Choroid"  
/dev\_stage="fetal and adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-E10"  
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-E10 is a subcloned cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGATCAGAG; lens, CGATTACCGA; eye anterior segment, AATGCCGAT; optic nerve, CCATTAGTG; retina, CCGCG; Retina foveal and Macular, GTCG; RPE and Choroid, ACCCA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

## ORIGIN

Query Match 41.6%; Score 694.4; DB 12; Length 712;  
Best Local Similarity 99.6%; Pred. No. 1.5e-99;  
Matches 706; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 667 CCGATGATTAACCTGTTCACTACAGCTGCACGTTCCAGAGCAATTCCTCCACTA 726  
Db 1 CCGATGATCACTGTTCACTACAGCTGCACGTTCCAGAGCAATTCCTCCACTA 60  
QY 727 ACTACCAATTAATCTGAGAGCATGTAAGTCCGTAAGTCACTCCGTAACCTGCTTTG 766  
Db 61 ACTACCAATTAATCTGAGAGCATGTAAGTCCGTAAGTCACTCCGTAACCTGCTTTG 120  
QY 787 CAGTCCAAAAACAAGCTTAGAAAAAGACTGCTCGGTGTGTATGTACAGTCACTGAT 846  
Db 121 CAGTCCAAAAACAAGCTTAGAAAAAGACTGCTCGGTGTGTATGTACAGTCACTGAT 180  
QY 847 GACCCACCATTCAGACAGGAGCATGTAAGTCCGTAAGTCACTCCAGTCACTGAT 906  
Db 181 GACCCACCATTCAGACAGGAGCATGTAAGTCCGTAAGTCACTCCAGTCACTGAT 240  
QY 907 TCCATATGTAATGTTACGAAACAAGACCTCCCTCGAGAGCTGAAAAATCCAGCTCT 966  
Db 241 TCCATATGTAATGTTACGAAACAAGACCTCCCTCGAGAGCTGAAAAATCCAGCTCT 300  
QY 967 ATGGATCCGATGCGCTTTATAGATCATTCAGACGTAATAGTTATCTTACGTTTGAG 1026  
Db 301 ATGGATCCGATGCGCTTTATAGATCATTCAGACGTAATAGTTATCTTACGTTTGAG 360  
QY 1027 AATGCAATTTGATGACTACATCACTGAGAAATTCGAGAGCCACTGAAACTGGGGGTA 1086  
Db 361 AATGCAATTTGATGACTACATCACTGAGAAATTCGAGAGCCACTGAAACTGGGGGTA 420

QY 1087 GTCCCTGATATATACGATATCCCGGAGCATCAGACTGGCTTCCAGTAACAAAGTCT 1146  
Db 421 GTCCCTGATATATACGATATCCCGGAGCATCAGACTGGCTTCCAGTAACAAAGTCT 480  
QY 1147 ATCTTTGATATCAAAATTTTCTACCCCGAGGAGTCAAGTTATCATCAGACCTGAT 1206  
Db 481 ATCTTTGATATCAAAATTTTCTACCCCGAGGAGTCAAGTTATCATCAGACCTGAT 540  
QY 1207 TCTGATGACATATGATATAGAGCCCTATGTAAGTAAGTAAGTGAAGTGAATCTTAC 1266  
Db 541 TCTGATGACATATGATATAGAGCCCTATGTAAGTAAGTAAGTGAAGTGAATCTTAC 600  
QY 1267 CAGCAGCTTTCAGACGCTCTCAGGAAACGAAATGGGAGTGCAGACGTCACAGAC 1326  
Db 601 CAGCAGCTTTCAGACGCTCTCAGGAAACGAAATGGGAGTGCAGACGTCACAGAC 660  
QY 1327 AATTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1375  
Db 661 AATTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 708

RESULT 7  
BX348736  
LOCUS BX348736 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens  
DEFINITION  
BX348736  
ACCESSION BX348736  
VERSION BX348736.1 GI:3037315  
KEYWORDS EST.  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4610.r For more information about this cluster, see  
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BA009ZG08\_CS00831\_1cluster=4610.r.  
Contact : Feng Liang Email : fliang@life.techn.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0BA009ZG08\_CS00831\_1.

## FEATURES

source

1..960  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D001YG24"  
/tissue\_type="NEUROBLASTOMA COT 50-NORMALIZED"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 41.1%; Score 686; DB 13; Length 960;  
Best Local Similarity 99.2%; Pred. No. 2.5e-98;  
Matches 700; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 940 CCGTACAGAGCTGAAAAATCCAGCTTATGATGATGATGATGATGATGATGATGATGAT 999  
Db 7 CCGTACAGAGCTGAAAAATCCAGCTTATGATGATGATGATGATGATGATGATGATGAT 66  
QY 1000 CAGTATAGTTATTCCTAGCTTTTGAAGTACGAGTTTGTGATCATCATCACTGAGAG 1059



DB 67 CAGTAAATCTTATCTAGCTTTAGAGATCAGTTTGTGTGACTACATCATCGAANAG 126  
QY 1060 TCTGTGAGGCGCACTGAAACTGGGGGTAGTCCCTGATATTCOGATCCCGCAGATACCA 1119  
DB 127 TTTCTGAGGCCCTGAAACTGGGGGTAGTCCCTGATATTCOGATCCCGCAGATACCA 126  
QY 1120 GACTGCTTCCAGTAACAAAAGTCTATCTTGTATGCAATTTCTCACCCCGAGGAA 1179  
DB 187 GACTGCTTCCAGTAACAAAAGTCTATCTTGTATGCAATTTCTCACCCCGAGGAA 246  
QY 1180 CTGGCAGGTATATAGACGACTGATTTCTGATGACAGATTGTGTGAGGCTTATGTGAA 1239  
DB 247 CTGGCAGGTATATAGACGACTGATTTCTGATGACAGATTGTGTGAGGCTTATGTGAA 306  
QY 1240 TGGAACTGAAAGGTGAGATCTCTAACACGACCTTTCAGAGCTCTCAGGGAACGGA 1299  
DB 307 TGGAACTGAAAGGTGAGATCTCTAACACGACCTTTCAGAGCTCTCAGGGAACGGA 366  
QY 1300 TGGGAGTGCAGAGAGCTCAACGAGCAATTCATTCATTCGATTTGATGTGTGTGC 1359  
DB 367 TGGGAGTGCAGAGAGCTCAACGAGCAATTCATTCATTCGATTTGATGTGTGTGC 426  
QY 1360 ACCAAGTGTGGGCTAATATCAGGCTTCAGGAAAAGGCTTACCAACCAAAAGTGGAG 1419  
DB 427 ACCAAGTGTGGGCTAATATCAGGCTTCAGGAAAAGGCTTACCAACCAAAAGTGGAG 486  
QY 1420 GCAAGAGATACCACTGATGTTGCCAGAGCCCAAGTGTGCTTCTCACTCCG 1479  
DB 487 GCAAGAGATACCACTGATGTTGCCAGAGCCCAAGTGTGCTTCTCACTCCG 546  
QY 1480 ACTCAACCTTGTAGCTTTTGGAGAGATGTGATTTCCAGCTTGAACATCCAGAAA 1539  
DB 547 ACTCAACCTTGTAGCTTTTGGAGAGATGTGATTTCCAGCTTGAACATCCAGAAA 606  
QY 1540 GAAGCCCAAGCACTAAGGTGGCTGTATGATGATCAAACTTTGATCTCAAGCTTT 1599  
DB 607 GAAGCCCAAGCACTAAGGTGGCTGTATGATGATCAAACTTTGATCTCAAGCTTT 666  
QY 1600 TGGGGCTTATGATTTCAAGGACTGATTTCAAAAAGATCAAGATGAA 1645  
DB 667 TGGGGCTTATGATTTCAAGGACTGATTTCAAAAAGATCAAGATGAA 711

RESULT 8  
LOCUS BG332068 687 bp mRNA linear EST 27-FEB-2001  
DEFINITION 602432466P1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:4549985 5',  
mRNA sequence.  
ACCESSION BG332068  
VERSION BG332068  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1998)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: DCTD/DRP/Gazdar  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LINC1242 row: d column: 18  
High quality sequence stop: 635.  
Location/Qualifiers  
1. 687

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4549985"  
/issue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_18"  
/note="Organ: lung; Vector: pOTB7; Site: 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

Query Match 33.6%; Score 561; DB 12; Length 687;  
Best Local Similarity 95.5%; Pred. No. 1,3e-78;  
Matches 654; Conservative 0; Mismatches 20; Indels 11; Gaps 7;

QY 681 GTTCAACTACACTGCCAGTTTCAGAGGCAATTCACCTTCCACTAATCCCAATCTT 740  
DB 2 GTTCAACTACACTGCCAGTTTCAGAGGCAATTCACCTTCCACTAATCCCAATCTT 61  
QY 741 GGAGAGCTTGAAGTCTGAAATCGATCCGATCTTCTTGTGAGTCCAAAACAA 800  
DB 62 GGAGAGCTTGAAGTCTGAAATCGATCCGATCTTCTTGTGAGTCCAAAACAA 121  
QY 801 GCTTGAAGAAAGACTTGCTCCGCTGTGTATGTATGATCAGTGCAGCTGACCCACATCAGA 860  
DB 122 GCTTGAAGAAAGACTTGCTCCGCTGTGTATGTATGATCAGTGCAGCTGACCCACATCAGA 181  
QY 861 CAGGACAGCTATGTTGGCGAGCTGATGACTTACATGAGGTGATTCCTATGTTGATG 920  
DB 182 CAGGACAGCTATGTTGGCGAGCTGATGACTTACATGAGGTGATTCCTATGTTGATG 241  
QY 921 TTTAGCAAAACAAAGCTCCCTCAGCAGCTGAAAAATCCAGCTCATGATGGATCCGATG 980  
DB 242 TTTAGCAAAACAAAGCTCCCTCAGCAGCTGAAAAATCCAGCTCATGATGGATCCGATG 301  
QY 981 CTTTATAGATCATTCGACAGTAAAGTTATCTTATGCTTTTGAAGATCAGCTTTGTGA 1040  
DB 302 CTTTATAGATCATTCGACAGTAAAGTTATCTTATGCTTTTGAAGATCAGCTTTGTGA 360  
QY 1041 TGACTACATCACTGAGAAAGTCTGAGGCGCACTGAAATCTGGGGTATGCCCTGATATTA 1100  
DB 361 TGACTACATCACTGAGAAAGTCTGAGGCGCACTGAAATCTGGGGTATGCCCTGATATTA 420  
QY 1101 CGGATCCCCCAGCATCAGAGCTGGCTTCCAAAGTAACAAAAGTCTATTTCTGATCAGA 1160  
DB 421 CGGATCCCCCAGCATCAGAGCTGGCTTCCAAAGTAACAAAAGTCTATTTCTGATCAGA 480  
QY 1161 ATTTTCTACCCCGAGGAACTGGCAAGTTATACATGACGACTGGATTCTGATGACAGATT 1220  
DB 481 ATTTTCTACCCCGAGGAACTGGCAAGTTATACATGACGACTGGATTCTGATGACAGATT 539  
QY 1221 GATAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1277  
DB 540 GATAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 599  
QY 1278 GACAGCTTCAGGG- AACGAAATGCGAGTGCAGACG- TCACACAGACAA--TTA 1331  
DB 600 AACAGCTTCAGGGCAACGAACTGGGGAGTGCAGACGCTCAACACAGACAAATTAC 659  
QY 1332 CATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1356  
DB 660 ATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 684

RESULT 9  
LOCUS CF748965 751 bp mRNA linear EST 10-OCT-2003

DEFINITION UI-M-HJ0-cmj-a-17-0-UI.r1 NIH\_BMAP\_HJ0 Mus musculus cDNA clone  
 IMAGE:30631600 5', mRNA sequence.  
 ACCESSION CF748965  
 VERSION CF748965.1 GI:37645310  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 751)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Dr. James Lin University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mouseetl.html>  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)  
 Seq primer: pyx-5.  
 Location/Qualifiers  
 1..751  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30631600"  
 /tissue\_type="Upper Head"  
 /dev\_stage="9.5 and 10.5 dpc"  
 /lab\_host="DH10B (T1 phase resistant)"  
 /clone\_lib="NIH\_BMAP\_HJ0"  
 /note="Organ: Head; Vector: pyx-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGACTGAAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN  
 Query Match 26.9%; Score 448.8; DB 14; Length 751;  
 Best Local Similarity 79.6%; Pred. No. 4.7e-61;  
 Matches 544; Conservative 0; Mismatches 133; Indels 6; Gaps 1;

QY 976 GATGGCTTTATAGATCATTTGCAAGATTAAGTTATCTTACCTTTGAGAAATGCACTT 1035  
 |||||  
 Db 1 GATGGCTTTACAGACTCATTCGCCAGTATTAAGTTCATCTGCGCTTTGAGAAATGCGATC 60  
 |||||  
 QY 1036 TGTGATGACTACATCACTGAGAGATTCGTGAGAGCCAGCAAGAACTGGGGGTAGTCCCTGTA 1095  
 |||||  
 Db 61 TGTGACGATTACATCAAGAGANGTTCTGAGACACATCTGGGGGTCTGCTGCCCTG 120  
 |||||  
 QY 1096 TATTACGATGCCCGAGCATCAAGATGGCTTCCAACTAACAAGTGTATTCTTTGTA 1155  
 |||||  
 Db 121 TATTACGATGCCCGAGCATCTGCCGATGGCTTCCAACTAACAAGTGTATTCTTTG 180  
 |||||  
 QY 1156 TCGAATTTTTCACCCCGAGGAATGGCAAGTTACATCAGACGACTGGATTCTGATGAC 1215  
 |||||  
 Db 181 TCGGAATTTCTTACCCCGAGAGCTGGGAGACTTATGAGAGCTGGAATTATGAGAT 240  
 |||||  
 QY 1216 AGATTATGAGGCCCTATGTAGAAATGGAAGTGAAGGTTGAGATCTTAACACGCACTT 1275  
 |||||

Db 241 GGGTTATGAGACCTATGTAGTGAAGCTAAAGGCAAGATCTCTTAACACGCGCTT 300  
 |||||  
 QY 1276 CTGACAGCTCTCAGGGAACGGAATGGGAGTGCAGAGCTCAACGAGCAATTAATC 1335  
 |||||  
 Db 301 CTGACAGCTCTCAATGAACGGGAGTGGGAGTGTGACGACTTAACGAGCAACTCAT 360  
 |||||  
 QY 1336 GATGATTTAGTGTATGTGTGTGACCAAGTGTGGCTTAATATCAGGCTTCAGAAAG 1395  
 |||||  
 Db 361 GACTGTTTGAAGTATGTGTGTGCGGAGGTTGGGCAACGATAGCTGCGAGGACG 420  
 |||||  
 QY 1396 GGCCTACACCCCAAAAGATGGAGGAGCAAGATACCCAGCTGAGTTGCCAGAGCCACA 1455  
 |||||  
 Db 421 GGCCTACACCCCAAAAGATGAAGGAGAGATTTAGCCACTGCATTCGCCGAGCTGCC 480  
 |||||  
 QY 1456 GTGTTGCTTTCTCACAC-----TCGGACTCCACCTTTGAGCTTTTGCAGAGATG 1509  
 |||||  
 Db 481 TTGTTACACCTTCTCATCTCCAGCTTCTCCGCCCTACGTGGAGATCGTTGCGAGAGCTG 540  
 |||||  
 QY 1510 TGGATTTCCAGTTTGAACATCCAGAAAGAGCCAGGACCTAAGTGGCTGTGAT 1569  
 |||||  
 Db 541 TGGCTACTAGCTTCCACAGCTCTAATAAGAGCCAGGACTGAGTGGCTGTGAT 600  
 |||||  
 QY 1570 AGGAATCAAACTTTTCACTCAAGAGTTTGGGCGCTAGTATCAAGACTGATTCAA 1629  
 |||||  
 Db 601 AGGAATCAAAATTTTTCATCTGAAGAGTTGCGGCTCTAGTATTAAAGCACTAATTTANA 660  
 |||||  
 QY 1630 AATGATCGAATGAAGAAACAGAA 1652  
 |||||  
 Db 661 ACGTACTGAGACTGACAGATTA 683  
 |||||

RESULT 10  
 Bf181631  
 Locus  
 DEFINITION Bf181631 945 bp mRNA linear EST 31-OCT-2000  
 601804381 NCI\_CGAP\_Mams Mus musculus cDNA clone IMAGE:4035584 5',  
 mRNA sequence.  
 Bf181631  
 Bf181631.1 GI:11059773  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 945)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LHM9309 row: x column: 09  
 High quality sequence stop: 645.  
 Location/Qualifiers  
 1..945  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4035584"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="7 months"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI\_CGAP\_Mams"  
 /note="Organ: mammary; Vector: pCMV-Sport6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Robin Humphreys,"

ORIGIN NIH

Query Match 26.6%; Score 444; DB 10; Length 945;  
 Best Local Similarity 81.9%; Pred. No. 2.2e-60;  
 Matches 560; Conservative 0; Mismatches 120; Indels 4; Gaps 4;

QY 699 GTTCAGCAGGCAATCCCACTTCCCACTTCACTAACCAGATCTGGAGAGCAATGAATCCT 758  
 DB 1 GTTCAGCAGGCAATCCCACTTCCCACTTCACTAACCAGATCTGGAGAGCAATGAATCCT 60

QY 759 GAAGTCACTCCCATPCCATGATCCCTTTCAGTCCCAAAAAAAGCTTGAAGAAACACTTGC 818  
 DB 61 GAAGTCACTCCCATPCCATGATCCCTTTCAGTCCCAAAAAAAGCTTGAAGAAACACTTGC 120

QY 819 TCCGCTGTGTATGTACAGTCACTGTGACCCACCATCAGACAGGACAGCACTATGTTGC 878  
 DB 121 -CCCATGTATATGTTCAGTCCGATCCGATCCACATCAGACAGGACAGCACTATGTTGC 179

QY 879 CGAGCTGATGATTTACATCAGAGTGGATTCCTATGAGTGAAGCTTGAAGAAACAGCT 938  
 DB 180 GAGGCTGATGAGTTCATGATGATGATTCCTATGAGTGAAGCTTGAAGAAACAGCT 239

QY 939 CCTCAGCAGCTGAAAAATCCAGCTCTATGATGATGCCAGTGGCTTTATAGATCATTGC 998  
 DB 240 TCTCAGCAGCTGAAAAATCCAGCTCTATGATGATGCCAGTGGCTTTATAGATCATTGC 239

QY 999 AAGATTAAGTTATCTTACATGCTTTTGAAGAGCACTTGTATGATGATCATCATCTAGAA 1058  
 DB 300 CCAATTAAGTTATCTTACATGCTTTTGAAGAGCACTTGTATGATGATCATCATCTAGAA 359

QY 1059 GTTCTGAGAGGCACTGAAATCTGGGGGTATGCTCTATATATATACGATCCCAAGCATCAG 1118  
 DB 360 GTTCTGAGAGGCACTGAAATCTGGGGGTATGCTCTATATATATACGATCCCAAGCATCAG 419

QY 1119 AATAGTCTCTCAAGTAAACAAAAGTCTATTTCTGTATCAGAAATTTCTCAACCCAGGGA 1178  
 DB 420 CCACTGCTCTCAAGTAAACAAAAGTCTATTTCTGTATCAGAAATTTCTCAACCCAGGGA 479

QY 1179 ACTGGCAATTCATCAGACGACGTGATTCATGATGAGCAAGTTGATAGGCTATGATGA 1238  
 DB 480 GCTGGCAATTCATCAGACGACGTGATTCATGATGAGCAAGTTGATAGGCTATGATGA 539

QY 1239 ATGGAAGCTGAAGGGGTGAGATCTCTACACAGCACTTCTCAGAGCTCTCAGGAAACGGA 1298  
 DB 540 GTGGAAGCTGAAGGGGTGAGATCTCTACACAGCACTTCTCAGAGCTCTCAGGAAACGGA 597

QY 1299 ATGGGAGTGCAGACGTCACACAGCAAGATTTACATGATGATGATGATGATGATG 1358  
 DB 598 GTGGGAGTGCAGACGTCACACAGCAAGATTTACATGATGATGATGATGATGATG 656

QY 1359 CACCAAGTGTGGCTAATATCAG 1382  
 DB 657 GCCGACGTGTGGCTAATATCAG 680

RESULT 11  
 BE751825 562 bp mRNA linear EST 25-APR-2001  
 LOCUS 204047 MARC 280V Bos taurus cDNA 5', mRNA sequence.  
 DEFINITION BE751825  
 ACCESSION BE751825  
 VERSION BE751825.1 GI:10165817  
 KEYWORDS EST.  
 SOURCE Bos taurus (cow)  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 1 (bases 1 to 562)  
 Smith,T.P.L., Grosse,W.M., Preking,B.A., Roberts,A.J., Stone,R.T.,  
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrner,K.,S.C.,  
 Bennett,G.L., Heaton,M.P., Jaegerfeld,W.W., Rohrer,G.A.,  
 Chitko-McKown,C.G., Perlea,G., Holt,I., Karanycheva,S., Liang,F.,

TITLE Quackenbush,J. and Keele,J.W.  
 JOURNAL Sequence evaluation of four pooled-tissue normalized bovine cDNA  
 MEDLINE libraries and construction of a gene index for cattle  
 PUBMED Genome Res. 11 (4), 626-630 (2001)  
 2118013  
 11282978

COMMENT Contact: Smith TPJ  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@mea1.marc.usda.gov  
 Single pass sequencing. Bases called and alt trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -mismatch 12 options.  
 PCR primers  
 FORWARD: AGGAAACAGCATGACCAT  
 BACKWARD: GTTTCCTCCATCAGAG  
 Plate: 44 row: G column: 22  
 Seq primer: ATTAGGAGACATATAG.  
 Location/Qualifiers  
 1..562  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 /tissue\_type="pooled"  
 /lab\_host="MDH10B"  
 /clone\_lib="MARC 280V"  
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
 library made from pooled tissue from testis, thymus,  
 semitendinosus muscle, longissimus muscle, pancreas,  
 adrenal, and endometrium."

ORIGIN

Query Match 25.3%; Score 422.6; DB 10; Length 562;  
 Best Local Similarity 85.5%; Pred. No. 7.5e-57;  
 Matches 483; Conservative 0; Mismatches 79; Indels 3; Gaps 1;

QY 948 GCTGAAAATCCAGCTCTATGATGCGGATGGCTTTATAGATCATTCAGAGTAA 1007  
 DB 1 GTTGGCAACCCGCTGATGAGACGCCGATGCTTTTACAGATCTCCCAAGTAA 60

QY 1008 GTTATCTTACCTTTTGAAGATGCGATTTGATGATGATGATGATGATGATGATG 1067  
 DB 61 GTTATCTTCTTCTTTGAGATGCGGTGCGATGATGATGATGATGATGATGATG 120

QY 1068 GCCATGAAATCTGGGGGTAGTCCCTGTATATTCAGATTCGCCAGCATCTCAGATGCGCT 1127  
 DB 121 ACCTGTAGGCTGGGGGTGCTCCTGTGTATGAGATCCCAAGCATCTCAGATGCGCT 180

QY 1128 TCCAGTAAACAAAGTGTATTTCTTGTATCAGATTTTCTACCCCAAGGAGTGGCAG 1187  
 DB 181 TCCAGTAAACAAAGTGTATTTCTTGTATCAGATTTTCTACCCCAAGGAGTGGCAG 240

QY 1188 TTACTCAGACGATGATTTGTATGACAGATTTGATGAGCCCTATGTATGATGAGT 1247  
 DB 241 CTACATCAGACGATGATTTGTATGACAGATTTGATGAGCCCTATGTATGATGAGT 300

QY 1248 GAAGGGTATGATCTCTAACCAGCATCTTGACAGCTCTCAGGGAAACGAAATGGGAGT 1307  
 DB 301 GAAGGGTATGATCTCTAACCAGCATCTTGACAGCTCTCAGGGAAACGAAATGGGAGT 360

QY 1308 GCAAGAGCTCAACGAGCAATTCATGATGATGATTTGAGTGTATGATGATGATGATG 1367  
 DB 361 GCAGAGAGCTTAAAGAGAGCACTATGATGATGATTTGAGTGTATGATGATGATGATG 420

QY 1368 GTGGGCTAATTCAGGGCTTCGAGAAAGGGCTTACCAACCCAAAGATGGAGGAGAGA 1427  
 DB 421 GTGGGCTAATTCAGGGCTTCGAGAAAGGGCTTACCAACCCAAAGATGGAGGAGAGT 480

QY 1428 TACCACCTGAGTGGCCAGAGCCCAAGATGTTGCTTTCTACCACTCGGATCGCAC 1487  
 DB 481 TACCACCTGAGTGGCCAGAGCCCAAGATGTTGCTTTCTACCACTCGGATCGCAC 537

**Qy** 1488 TTGAGCTCTTTGCGAGAGATGTGG 1512  
**Db** 538 TCGGCGCTCTTTGCGAGAGATGTGG 562  
**RESULT 12**  
**LOCUS** BU261379 849 bp mRNA linear EST 26-NOV-2002  
**DEFINITION** 603374248F1 CSEQCHN51 Gallus gallus cDNA clone CHEST286012 5', mRNA sequence  
**ACCESSION** BU261379  
**VERSION** BU261379.1 GI:25528988  
**KEYWORDS** EST  
**SOURCE** Gallus gallus (chicken)  
**ORGANISM** Gallus gallus  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
**AUTHORS** 1 (bases 1 to 849)  
 Bortman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
**TITLE** A Comprehensive Collection of Chicken cDNAs  
**JOURNAL** Curr. Biol. 12 (22), 1965-1969 (2002)  
**MEDLINE** 22335534  
**PUBMED** 12445392  
**COMMENT** Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.  
**FEATURES**  
**source**  
 Location/Qualifiers  
 1..849  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST286012"  
 /dev\_stage="22"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQCHN51"  
 /note="Organ: 1limb; Vector: Bluescript II KS(+); Site\_1: EcoRI; Site\_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the Bluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996) 791, except that a significantly longer reannealing hybridization was used."  
**ORIGIN**  
**Query Match** 24.9%; Score 415.2; DB 13; Length 849;  
**Best Local Similarity** 70.0%; Pred. No. 7.9e-56;  
**Matches** 588; Conservative 0; Mismatches 248; Indels 4; Gaps 2;  
**Qy** 766 CTCGATACCTAGTCTTGGACATCAAAACAAGCTTAGAAAAGACTGTCCGCTG 825  
**Db** 3 CTGAGACATATGATCCAGTGCAGATGAAGAACACCTGAGGAACGGCTTGACACACTT 62  
**Qy** 826 GTGTATGTACAGTCAAGCTGTGACCCACCATCAGACAGGAGAGCTTATGTTCCGAGCTG 885  
**Db** 63 GTGTATGTGACAGTCAAGCTGTGATGCTCTTCTGACCGGAGACGTATGTACGTGAGTTG 122  
**Qy** 886 ATGACTTACATGACAGGTGATTCCTATGTGTAATGTTTACGAAACAAAGACCTCCCTCAG 945

**Db** 123 ATGTCCACATGTGAAGTACTTACGGAGAAATGTGATACAGAGACCTTCCCTCAG 182  
**Qy** 946 CAGCTGAAAATCCAGCCTTATGATGCCGATGCTTTTATAGATCATTTGACACAGTAT 1005  
**Db** 183 CATTCGAAATTCATCTGCATGATGATGAGAACTTCTTAAATATCTGACACAGTAC 242  
**Qy** 1006 AAGTTATCTAGCTTTTGAAGATGACAGTTTGTATGATCAATCATCTGAGAAAGTTCTGG 1065  
**Db** 243 AAGTTATCTAGCTTTTGAAGATGATCTGATGATATATCATCTGAAAAAATCTCTGG 302  
**Qy** 1066 AGGCCATGAAACTGGGGGTAGTCCCTGATATTAAGGATCCCCGACATCAGAGCTGG 1125  
**Db** 303 CGGCTCTGATGCTGGGAGTGGTACCTGTGATCTTGTCTCCAGCATTTATGCACTGG 382  
**Qy** 1126 CTTCAGATTAACAAAAGTGTATTTCTGTATCAGAAATTTCTACCCGAGGAACTGGCA 1185  
**Db** 363 CTTCACAGACAAAGAGTGCATCTGATCTGATCTAGTTTTCACACCTCGAGAGCTGGCC 422  
**Qy** 1186 AGTTACATCAGACAGCTGATTTCTGATGACAGATTTGATAGGCTTATGATGAGAG 1245  
**Db** 423 CCTATATCAAAACCTGTGATCAAAATGACAGAGATGAGGCTTACTGTGAATGAAA 482  
**Qy** 1246 CTGAAGGTGAGATCTCTAACCCAGCACTTCTGACAGCTCTCAGAGGAAAGGAATGGGGA 1305  
**Db** 483 CTGAAGAGAGACATTTCCATCCAAAGCTGCTTACAGCATGAAAGAACCAATGGGGA 542  
**Qy** 1306 GTGCAGACGTCACACAGACAAATTAATCATGATCATTTGATGATGATGTCACCAAG 1365  
**Db** 543 GTGCAGAGATGTCACACAGACAAATTAATTAATGACATTTGATGATGATGATGATGAC 602  
**Qy** 1366 GTGTGGGCTAATATGAGGCTTCAGAAAAGGCTTACACCCAAAGATGAGGAGAA 1425  
**Db** 603 GTGTGGGAAACATATGAGAGAAAGAAAGAGAGATGCTGCCACAGAGTGAATGCTCAG 662  
**Qy** 1426 GATACCAACTGAGTTGGCCAGAGCCACAGCTGTTTCTTACCACTCCGACTCCA 1485  
**Db** 663 GTTAACCATCTGAATGCTCCAAACCTGAGGCTTCTGGATTCCTCCTC---TTCAAAATCCC 719  
**Qy** 1486 CTTTGACTCTTTTCCGAGATGTGATTTCCAGCTTTTAAACAATCCAGAAAGAGCC 1545  
**Db** 720 GGCAGACTCTCTCCAGAGATGTGATGACAGATTTTGAAGCATCAAGAGAGAGCC 779  
**Qy** 1546 CAGGACTAAGGTGGCTGTGATAGAAATCAAACTTTTATCTCAAGAGTTTGGGGC 1605  
**Db** 780 TGGGGCTGAGGAGAGCTGTGAAAGAGAA-CAGAAATTTCAACTCAAGAAATTTGATGC 838  
**RESULT 13**  
**LOCUS** BM090273 523 bp mRNA linear EST 19-NOV-2001  
**DEFINITION** 505452 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.  
**ACCESSION** BM090273  
**VERSION** BM090273.1 GI:17000901  
**KEYWORDS** EST.  
**SOURCE** Bos taurus (cow)  
**ORGANISM** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
**REFERENCE** 1 (bases 1 to 523)  
 Smith, T.P.L., Grose, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Caeas, E., Wray, D.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chalko-McKown, C.G., Perlea, G., Holt, I., Katanycheva, S., Liang, F., Quackenbush, J. and Keesle, J.W.  
**TITLE** Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle  
**JOURNAL** Genome Res. 11 (4), 626-630 (2001)  
**MEDLINE** 21180013  
**PUBMED** 11282978  
**COMMENT** Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -mismatch 12 options.

PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTCCAGCTACGACG  
Plate: 97 row: E column: 4  
Seq primer: ATTGAGTACACTATAG.

FEATURES  
source  
1..523  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6, Site\_1: NotI; Site\_2: SalI; library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."

# ORIGIN

Query Match 23.4%; Score 390; DB 12; Length 523;  
Best Local Similarity 85.2%; Pred. No. 1e-51; Matches 448; Conservative 0; Mismatches 75; Indels 3; Gaps 1;

1066 AGGCACTGAACTGGGGGTAGTCCCTGATATTAACGATCCCGACATACAGACTGG 1125  
1 AACACCTGTAACCTGGGGGGCGCTCCCTGTGATTAACGATCCCGACATACGAGACTGG 60  
1126 CTTCCAGTAAACAAAGTCTATTCTTGTATCAGATTTTCTACCCGAGGAACTGGCA 1185  
61 CTTCCAGTAAAGAGTCTATTCTTGTATCAGATTTTCTACCCGAGGAGCTGGCG 120  
1186 AGTTCATCAGACGACTGATCTGATGACAGATTTGATGAGGCTATGTAGATGGAAG 1245  
121 AGTTCATCAGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 180  
1246 GTGAAGGAGTGAATCTTAACCAAGGACTTTGACAGCTTCAGGGAACGGAATGGGGA 1305  
181 CTGAAGGAGGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 240  
1306 GTGCAAGAGTGAACCAAGGACATTAACATGATGATGATGATGATGATGATGATGAT 1365  
241 GTGCAAGAGTGAACCAAGGACATTAACATGATGATGATGATGATGATGATGATGAT 300  
1366 GTGTGGGCTAATATAGGCTTCAAGAAAGGCTTACCAACCAAAAGATGGAGCAGAA 1425  
301 GTGTGGGCTAATATAGGCTTCAAGAAAGGCTTACCAACCAAAAGATGGAGCAGAA 360  
1426 GATTAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1485  
361 GTTAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 417  
1486 CATTGAGCTCTTTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1445  
418 CATTGAGCTCTTTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 477  
1546 CAGGCACTAAGGTGGCTGGTGTAGATGAATCAAACTTTTCAATCTC 1591  
478 CAGGCACTAAGGTGGCTGGTGTAGATGAATCAAACTTTTCAATCTC 523

RESULT 14  
BU449301  
LOCUS BU449301 728 bp mRNA linear EST 29-NOV-2002  
DEFINITION 603764226f1 CSEORBN13 Gallus gallus cDNA clone CHEST687d8 5', mRNA  
sequence.  
ACCESSION BU449301

VERSION BU449301.1 GI:25938612  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 728)  
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
A Comprehensive Collection of Chicken CDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)  
22335534  
MEDLINE  
12445392  
PUBMED

## COMMENT

Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 10D, UK  
Tel: 01612068930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.

## FEATURES

source

1..728  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="layer"  
/db\_xref="taxon:9031"  
/clone="CHEST687d8"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="CSEORBN13"  
/note="Organ: ovary; Vector: pBluescript II KS(+); Site\_1:  
EcoRI; Site\_2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was blunt-ended, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

## ORIGIN

Query Match 23.3%; Score 388.8; DB 13; Length 728;  
Best Local Similarity 72.7%; Pred. No. 1.2e-51; Matches 529; Conservative 0; Mismatches 197; Indels 2; Gaps 2;

635 CCCCAGAAAACAAATTAAGTCTTTTATTAACGAGATTAACCTGTTCACTACACTG 694  
1 CACCAAAAACAACTACCAAACTTTTCAATGAAACGATACCTTATTCAACCACTG 60  
695 CCACGTCAGCAGGATTCACCACTTGCACATTAACCAATCTTGAGAGGATTTGAAG 754  
61 CAACCTTAGCGCGCATTTCTACCTACCGGTGACCACTGATGATGATGATGATGATG 120  
755 TCCTGAGTCTCTTGAACATATGATCCAGTGACAGATGAAGAAACAGCTTGAAGAGCG 180  
121 TCCTGAGTCTCTTGAACATATGATCCAGTGACAGATGAAGAAACAGCTTGAAGAGCG 180  
815 TTGTCCTGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 874  
181 TTGTCCTGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 240  
875 TTGTCCTGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 934  
241 TACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300

QY 935 ACCTCCCTCAGCAGTGAATAATCCAGCCTCTATGATGCCAGTCTTTTATAGATCA 994  
 Db 301 ACCTCCTCAGCATCTCAGAAATCCATCTGCCATGGATGATGGAACTTCTATATAATAC 360  
 QY 995 TTGCACAGTATAAGTTATCTTACGCTTTGAGAAATGCAAGTTGTATGACTACATCACTG 1054  
 Db 361 TGGACAGTACAAAGTTCTTCTTTGAAAATGCTATGCTGTAAGATTATATCACTG 420  
 QY 1055 AGAAGTCTGAGGACCACTGAAACCTGGGGGTAGTCCCTGTATATTAACGATCCCAACA 1114  
 Db 421 AAAAATCTGCGCGCTCTGATGCTGGAGTGGTACTGTACTTGGTTCTCCACAGA 480  
 QY 1115 TCACAGCTGCTTCCAGATACAAAAGTCTATTCTGTATCAGAAATTTCTCAACCCA 1174  
 Db 481 TTATCGACTGGCTTCCAGACACAAAGTGCATCTTGGTATCTAGTTTTCACACCTTC 540  
 QY 1175 GGGAACTGGCAAGTTATCATGACGACATGCTGATTCGATGACAGATTGTATGAGGCTATG 1234  
 Db 541 GAGAGCTGGCCCTTATATCAAAACCTGGATCAAAATGACCGAGATGAGGCTATCC 600  
 QY 1235 TAGAATGGAAGCTGAA-GGGTGAGATCTCTAAACAGCAGCTTCTGACAGCTCTCAAGGAA 1293  
 Db 601 TGAATGGAAGCTGAAAGAGAGACATTTCCAAATCCAAAGCTCTTACAGCAATGAAGAA 660  
 QY 1294 CGGAATGGGAGTGCAGAACGTCACACGACGCAAT-TACATCGATGCAATTTGAGTGTAT 1352  
 Db 661 CGGAATGGGAGTGCAGAAATGTCACCTGAGCAATCTATATCGACACATTTGAGTGTAT 720  
 QY 1353 GGTGTGCA 1360  
 Db 721 GGTGTGTA 728

RESULT 15  
 BX873167 839 bp mRNA linear EST 17-DEC-2003  
 LOCUS BX873167 AGENAE Rainbow trout multi-tissues-normalized (tcbk)  
 DEFINITION Oncorhynchus mykiss CDNA clone tcbk0017c.p.23 5prim, mRNA sequence.  
 BX873167  
 VERSION BX873167.1 GI:40001712  
 KEYWORDS EST.  
 SOURCE Oncorhynchus mykiss (rainbow trout)  
 ORGANISM Oncorhynchus mykiss  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 Protecanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 1 (bases 1 to 839)  
 GOVERNOR, M., Guiguen, Y. and Le Gac, F.  
 REFERENCES Construction and primary characterization of normalized cDNA  
 libraries in rainbow trout, *Oncorhynchus mykiss*  
 Unpublished (2003)  
 JOURNAL Contact: Guiguen Y  
 COMMENT INRA - SCRIBE  
 Campus de Beaulieu, RENNES cedex, 35042, France  
 Tel: 02.23.48.50.09  
 Fax: 02.23.48.50.20  
 Email: Yann.Guiguen@beaulieu.rennes.inra.fr  
 Sequence cleaned of vector, adaptor and repetitions. Contact us  
 at sigenasupport@jouy.inra.fr to obtain the chromatogram of this  
 sequence.  
 Plate: 0017 row: P column: 23  
 Seq primer: M13R.  
 Location/Qualifiers  
 1. 839  
 /organism="Oncorhynchus mykiss"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8022"  
 /clone="tcbk0017c.p.23"  
 /tissue\_type="multi-tissues"  
 /dev\_stage="from embryos to adults"  
 /lab\_host="DH10B"  
 /clone\_lib="AGENAE Rainbow trout multi-tissues-normalized  
 (tcbk)"

/note="Vector: pT73D-pac; Rainbow trout  
 multi-tissues-normalized + 2 subtractions; Clone  
 distribution: AGENAE Resource centre, Francois Piumi,  
 Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et  
 Etude du genome (UREG), Domaine de Vilvert, 78352,  
 Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33  
 (0) 1.34.65.22.73"

## ORIGIN

Query Match 21.0%; Score 350.6; DB 13; Length 839;

Best Local Similarity 65.0%; Pred. No. 1.1e-45;

Matches 554; Conservative 0; Mismatches 285; Indels 3; Gaps 1;

QY 732 CCAATAGTGGAGAGCATGGAAGTCTGAAATCACTCCGATACCTAGTCTTTGAGATC 791  
 Db 9 CCAGACCTGGGAACTGGAGTCTGACCTCCCAAGCTACCTGGTGTCCCTCAACCA 68  
 QY 792 CAAAACAGCTTAGAAAAGACTGTGCTCGCTGTGTATGTACGTACAGCTGACCC 851  
 Db 69 GAAAGACAGCTGAGGAAACACCTAGCCCGGTGTAGCTCAGACGACGACCC 128  
 QY 852 ACCATCAGACAGGACAGCTATGTTGCGAGCTGATGACTTACATCGAGGTGATTCCTA 911  
 Db 129 TCCGTGGACCGGGATGCTTATGTGATGAGTTATGAGCAATCAGGGTTGATTCCTA 188  
 QY 912 TGGTGAATGTTTACGAAACAAAGACTCTCTCAGACAGCTGAAAATCAGACTCTATGGA 971  
 Db 189 TGGACAAATGCCCTCCACAAACAAAGACTGCTCTCCATCTAGAGGACTCGATTGCCATGA 248  
 QY 972 TSCCGATGCTTTTATAGATCATTTGCAAGATATATGTTATTCCTAGCTTTGAGATGC 1031  
 Db 249 GAGACAGGCTTTCTTACAGATCTCGGCCAGTAAATTCATTTGCGCTTTGAGAACGC 308  
 QY 1032 AGTTGTATGATCTATCATCACTGAGAAAGTTCTGAGGSCCACTGAAATCGGGGTATGCC 1091  
 Db 309 CCTCTGTATCAGACTATCAATCACCAGAAACTATGAGGCTCTCAAGCTTGGGTGGTTCC 368  
 QY 1092 TGTATATTAAGGATCCCCAGCATCAACAGCTGGCTTCCAAAGTAAACAAAGTCTATTC 1151  
 Db 369 TGTGATTAACGCTGCTCCCAATGTTTGGCAATGGCTCCAGCAACAGTGTGCGGTGT 428  
 QY 1152 TGTATCAGAAATTTCTCAACCCAGGAGAACTGCAAGTTACATCAGACGATGATTCGA 1211  
 Db 429 AGTCACCCAGACGAGGCCCTTGAATAAATGCTGTATCTGAAGAGATGATAGAGA 488  
 QY 1212 TGCACGATTTGATAGGCTTGTGTAATGGAAGCTGAAGGTGAGATCTCTAACAGCG 1271  
 Db 489 CGACGAGAGATGATGCAATTAATCTGAGGTGAAGTTAAACGTGAAGTCTCAACCTGGA 548  
 QY 1272 ACTTGTGACAGCTCTCAGAGGAACCGAAATGGGAGTCAAGAGCTCAACGAGACAATTA 1331  
 Db 549 GTTGTGTAAGAGAGCTGAAGAAAGCCCATGGGGTGTCAAGATCTGACAGAGAACTA 608  
 QY 1332 CATCATCATTTTGAAGTATATGTGTGACCAAGGTGTGGCTATATTCAGGCTTAGGA 1391  
 Db 609 CATATATGTGTTGATGATCATGATGTGTGACAGAGGTGTGGAGAACTCAGACAGAGAA 668  
 QY 1392 AAAGGCTTTACACCCAAAGATGGAGGAGGAGAAATACCCACTGAGTGTCCAGAGCC 1451  
 Db 669 AAAGGCTTTGTTCTTAAACCTTGCAAGAGAGAGAAACCACTTAAAGTGTCACTTCC 728  
 QY 1452 CACAGTGTGCTTTTTCACACTCCGAGCTCCACTTTAGCTCTTTGGAGAGATGTG 1511  
 Db 729 GAGGATATTTGACTTGTTCAGGCC--CTAGAAAGCGACATCTTTGGAGAAATTTG 785  
 QY 1512 GATTTCAGCTTTGAAACATCCAGAAAGAGCCAGCACT 1553  
 Db 786 GACGCGCAGTTACGAGCAGTCCAAAGAAAGAGCTTAGGGCAATT 827

Search completed: September 13, 2004, 02:17:58

Job time : 3073.12 secs



QY 119 TYLHHMTKAFLEFGTDFNIDSLPLPRKAHDMVAFHESPKNKYLFHKFVITLFFNYTA 178  
DB 61 TYLHHMTKAFLEFGTDFNIDSLPLPRKAHDMVAFHESPKNKYLFHKFVITLFFNYTA 120  
QY 179 TFSRSHLPLTTOYLESIEVLKSRVLVLOSKNLRRLAPLVYVQSDCPSPDRSYV 238  
DB 121 TFSRSHLPLTTOYLESIEVLKSRVLVLOSKNLRRLAPLVYVQSDCPSPDRSYV 180  
QY 239 REMMTYIEVDSYGECLRNKDLPOQKNPASMADGFYRIIAQYKFIILAFENAVCDYITE 298  
DB 181 REMMTYIEVDSYGECLRNKDLPOQKNPASMADGFYRIIAQYKFIILAFENAVCDYITE 240  
QY 299 KFMRLPLKGVPPYVGSPTITDMLPSNKSALIVSFSHPRELASTIRLSDDLRYEAYV 358  
DB 241 KFMRLPLKGVPPYVGSPTITDMLPSNKSALIVSFSHPRELASTIRLSDDLRYEAYV 300  
QY 359 EMTLKGELISNORLLTALREKMGVQDVADNYIDAFCMCVCTKWANIRLOEKGLPPKRW 418  
DB 301 EMTLKGELISNORLLTALREKMGVQDVADNYIDAFCMCVCTKWANIRLOEKGLPPKRW 360  
QY 419 EADDTMLSCPEPTVFAFSPLRTPLPSLSREKMWISSFEOSKKEAQLRWLVDRNQNFSQOE 478  
DB 361 EADDTMLSCPEPTVFAFSPLRTPLPSLSREKMWISSFEOSKKEAQLRWLVDRNQNFSQOE 420  
QY 479 FMGLVFRD 486  
DB 421 FMGLVFRD 428

RESULT 2  
QYREB  
ID QYREB PRELIMINARY; PRT: 462 AA.  
AC QYREB;  
DT 01-OCT-2003 (TREMBlrel. 25, Created)  
DT 01-OCT-2003 (TREMBlrel. 25, last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)  
DE Putative alpha 1,3-fucosyltransferase.  
GN FUD10.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Germot A., Dupuy F., Lortol C., Julien R., Matrah A.;  
RT "genomic organization and expression profile of putative alpha 1,3-  
RT fucosyltransferase genes, fuc10 and fuc11, in Bos taurus."  
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY344579; AA002331.1; -  
KM Glycosyltransferase; Transferase.  
SQ SEQUENCE 462 AA; 53884 MW; 0DD0D79FCC880F35 CRC64;

Query Match 80.7%; Score 2119.5; DB 6; Length 462;  
Best Local Similarity 86.1%; Pred. No. 2.6e-169;  
Matches 390; Conservative 25; Mismatches 37; Indels 1; Gaps 1;

QY 34 EWMVELGKFERKEKSSSLQDGHTKMEAPTHLNSFLKKEGLTFNRKRWELDSYPIMLW 93  
DB 11 QYVVELGKFERKEKSSSLQDGHTKMEAPTHLNSFLKKEGLTFNRKRWELDSYPIMLW 70  
QY 94 WSPUTGEGRIGGCGACDCEFTTINRTYLHHMTKAFLEFGTDFNIDSLPLPRKAHDMVA 153  
DB 71 WSPUTGEGRIGGCGACDCEFTTINRTYLHHMTKAFLEFGTDFNIDSLPLPRKAHDMVA 130  
QY 154 FHESPPKNKYLFHKFVITLFFNYTATFSRSHLPLTTOYLESIEVLKSRVLVLOSKNK 213  
DB 131 FHESPPKNKYLFHKFVITLFFNYTATFSRSHLPLTTOYLESIEVLKSRVLVLOSKNK 190  
QY 214 LKRLAPLVYVQSDCPSPDRSYVREMTYIEVDSYGECLRNKDLPOQKNPASMADG 273  
DB 191 LKRLAPLVYVQSDCPSPDRSYVREMTYIEVDSYGECLRNKDLPOQKNPASMADG 250

QY 274 FYRIIAQYKFIILAFENAVCDYITEKFMRLKGVVYVYGSPTITDMLPSNKSALIVSE 333  
DB 251 FYRIIAQYKFIILAFENAVCDYITEKFMRLKGVVYVYGSPTITDMLPSNKSALIVSE 310  
QY 334 FSHPRELASYIRLSDDLRYEAYVEWKLKGELISNORLLTALREKMGVQDVADNYIDA 393  
DB 311 FSHPRELASYIRLSDDLRYEAYVEWKLKGELISNORLLTALREKMGVQDVADNYIDA 370  
QY 394 FECWCTKWANIRLOEKGLPPKKEAEADTHLSCPEPTVFAFSPLRTPLPSLSIREMWS 453  
DB 371 FECWCTKWANIRLOEKGLPPKKEAEADTHLSCPEPTVFAFSPLRTPLPSLSIREMWS 429  
QY 454 FEOSKKEAQLRWLVDRNQNFSQOEFGLVFRD 486  
DB 430 FEOSKKEAQLRWLVDRNQNFSQOEFGLVFRD 462

RESULT 3  
Q8IVJ3  
ID Q8IVJ3 PRELIMINARY; PRT: 419 AA.  
AC Q8IVJ3;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, last annotation update)  
DE Putative alpha 1,3-fucosyl transferase.  
GN FUD10.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Candelier U.J., Martinez-Duncker I., Oriol R., Mollicone R.;  
RT "Cloning and genomic organization of a new alpha1-fucosyltransferase  
RT FUD10."  
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ512465; CAD54669.1; -  
DR GO: GO:0016020; C:membrane; IEA.  
DR GO: GO:0008417; F:fucosyltransferase activity; IEA.  
DR GO: GO:0016740; F:transferase activity; IEA.  
DR GO: GO:0006486; F:protein amino acid glycosylation; IEA.  
DR InterPro: IPR01503; Glyco\_trans\_10.  
DR Pfam: PF00852; Glyco\_transf\_10; I.  
KM Transferase.  
SQ SEQUENCE 419 AA; 48854 MW; 6B17FF8315B53B5 CRC64;

Query Match 77.7%; Score 2040; DB 4; Length 419;  
Best Local Similarity 97.2%; Pred. No. 1.1e-162;  
Matches 380; Conservative 2; Mismatches 3; Indels 6; Gaps 1;

QY 34 EWMVELGKFERKEKSSSLQDGHTKMEAPTHLNSFLKKEGLTFNRKRWELDSYPIMLW 93  
DB 27 QYVVELGKFERKEKSSSLQDGHTKMEAPTHLNSFLKKEGLTFNRKRWELDSYPIMLW 86  
QY 94 WSPUTGEGRIGGCGACDCEFTTINRTYLHHMTKAFLEFGTDFNIDSLPLPRKAHDMVA 153  
DB 87 WSPUTGEGRIGGCGACDCEFTTINRTYLHHMTKAFLEFGTDFNIDSLPLPRKAHDMVA 146  
QY 154 FHESPPKNKYLFHKFVITLFFNYTATFSRSHLPLTTOYLESIEVLKSRVLVLOSKNK 213  
DB 147 FHESPPKNKYLFHKFVITLFFNYTATFSRSHLPLTTOYLESIEVLKSRVLVLOSKNK 206  
QY 214 LKRLAPLVYVQSDCPSPDRSYVREMTYIEVDSYGECLRNKDLPOQKNPASMADG 273  
DB 207 LKRLAPLVYVQSDCPSPDRSYVREMTYIEVDSYGECLRNKDLPOQKNPASMADG 266  
QY 274 FYRIIAQYKFIILAFENAVCDYITEKFMRLKGVVYVYGSPTITDMLPSNKSALIVSE 333  
DB 267 FYRIIAQYKFIILAFENAVCDYITEKFMRLKGVVYVYGSPTITDMLPSNKSALIVSE 326  
QY 334 FSHPRELASYIRLSDDLRYEAYVEWKLKGELISNORLLTALREKMGVQDVADNYIDA 393



Db 327 FSHPRELASIYRRLDSDDLRYEAYEWKLGKGISNORLLTALREKRWGVQDVNDNYIDA 386  
 QY 394 FECWCTKVMANIRLOEK-----GLPPKRW 418  
 Db 387 FECWCTKVMANIRLOEKVSKSGVIEPAGW 417

## RESULT 4

081V17 PRELIMINARY; PRT; 391 AA.  
 AC 081V17;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 GN Putative alpha3-fucosyltransferase.  
 DE FUT10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Candelier J.U., Martinez-Duncker I., Oriol R., Mollicone R.;  
 RT "Cloning expression and genomic organization of two new human alpha3-  
 fucosyltransferases (FUT10 and FUT11).";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ053836; C:membrane; IEA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008417; F:fucosyltransferase activity; IEA.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.  
 DR InterPro; IPR01503; Glyco\_transf\_10.  
 DR Pfam; PF00852; Glyco\_transf\_10; I.  
 DR Transferase; Glycosyltransferase.  
 SQ SEQUENCE 391 AA; 45696 MW; 332C1B4FDCDCA565 CRC64;

Query Match 77.5%; Score 2034; DB 4; Length 391;  
 Best Local Similarity 97.4%; Pred. No. 3,1e-162;

Matches 379; Conservative 1; Mismatches 3; Indels 6; Gaps 1;

QY 36 VWELEKFERKEPKSSLDGGHTKKEBAPTHLNSFLKKEGLTFNKRKWELEDSYPIIMWMS 95  
 Db 1 VWELEKFERKEPKSSLDGGHTKKEBAPTHLNSFLKKEGLTFNKRKWELEDSYPIIMWMS 60  
 QY 96 PLTGTGRLGGCGACACFTTNRITLHHMTKAFLEFGTDPNIDSLPLPKAHHDMAVFI 155  
 Db 61 PLTGTGRLGGCGACACFTTNRITLHHMTKAFLEFGTDPNIDSLPLPKAHHDMAVFI 120  
 QY 156 EESPKNNKYLFRKPYITLFNTATFSRSHSLPLTQYLESIEVLKSLRYLVPLQSKNKL 215  
 Db 121 EESPKNNKYLFRKPYITLFNTATFSRSHSLPLTQYLESIEVLKSLRYLVPLQSKNKL 180  
 QY 216 KKLAPLVYQSCDPPSDRSYVRELMTYIEVDSGELRNKQDLPQOLKNPASMDADGFY 275  
 Db 181 KKLAPLVYQSCDPPSDRSYVRELMTYIEVDSGELRNKQDLPQOLKNPASMDADGFY 240  
 QY 276 RIIAQYKFLAENAVCDYITEKFWRLKLGVPVYVYGSPTIDWLPSNKSALIVSEFS 335  
 Db 241 RIIAQYKFLAENAVCDYITEKFWRLKLGVPVYVYGSPTIDWLPSNKSALIVSEFS 300  
 QY 336 HPEELASVYRRLDSDDLRYEAYEWKLGKGISNORLLTALREKRWGVQDVNDNYIDA 395  
 Db 301 HPEELASVYRRLDSDDLRYEAYEWKLGKGISNORLLTALREKRWGVQDVNDNYIDA 360  
 QY 396 CWCCTKVMANIRLOEK-----GLPPKRW 418  
 Db 361 CWCCTKVMANIRLOEKVSKSGVIEPAGW 389

RESULT 5  
 08K0S3 PRELIMINARY; PRT; 436 AA.  
 AC 08K0S3;

DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Putative fucosyltransferase.  
 GN FUT10.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RT TISSUE-Regina;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC030474; AAH0474.1; .  
 DR MGD; MGI:2384748; Fuc10.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008417; F:fucosyltransferase activity; IEA.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.  
 DR InterPro; IPR01503; Glyco\_transf\_10.  
 DR Pfam; PF00852; Glyco\_transf\_10; I.  
 DR Transferase; Glycosyltransferase.  
 SQ SEQUENCE 436 AA; 50385 MW; 16A655F1027F8484 CRC64;

Query Match 66.1%; Score 1734.5; DB 11; Length 436;  
 Best Local Similarity 80.8%; Pred. No. 4.9e-137;

Matches 316; Conservative 34; Mismatches 38; Indels 3; Gaps 1;

QY 34 EVMWELGFERKEPKSSLDGGHTKKEBAPTHLNSFLKKEGLTFNKRKWELEDSYPIIMW 93  
 Db 27 QVMWELGFERKEPKSSLDGGHTKKEBAPTHLNSFLKKEGLTFNKRKWELEDSYPIIMW 86  
 QY 94 WSPLTGTRIGCGACACFTTNRITLHHMTKAFLEFGTDPNIDSLPLPKAHHDMAV 153  
 Db 87 WSPLTGTRIGCGACACFTTNRITLHHMTKAFLEFGTDPNIDSLPLPKAHHDMAV 146  
 QY 154 FHEESPKNNKYLFRKPYITLFNTATFSRSHSLPLTQYLESIEVLKSLRYLVPLQSKN 213  
 Db 147 FHEESPKNNKYLFRKPYITLFNTATFSRSHSLPLTQYLESIEVLKSLRYLVPLQSKN 206  
 QY 214 LRKLAPLVYQSCDPPSDRSYVRELMTYIEVDSGELRNKQDLPQOLKNPASMDADG 273  
 Db 207 LRKLAPLVYQSCDPPSDRSYVRELMTYIEVDSGELRNKQDLPQOLKNPASMDADG 266  
 QY 274 FYRIIAQYKFLAENAVCDYITEKFWRLKLGVPVYVYGSPTIDWLPSNKSALIVSE 333  
 Db 267 FYRIIAQYKFLAENAVCDYITEKFWRLKLGVPVYVYGSPTIDWLPSNKSALIVSE 326  
 QY 334 FSHPRELASVYRRLDSDDLRYEAYEWKLGKGISNORLLTALREKRWGVQDVNDNYIDA 393  
 Db 327 FSHPRELASVYRRLDSDDLRYEAYEWKLGKGISNORLLTALREKRWGVQDVNDNYIDA 386  
 QY 394 FECWCTKVMANIRLOEKGLPPKWEADTH 424  
 Db 387 FECWCTKVMANIRLOEK--VSEKSGGWH 414

## RESULT 6

08AMB5 PRELIMINARY; PRT; 439 AA.  
 AC 08AMB5;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 GN Putative alpha3-fucosyltransferase.  
 DE FUT10.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NC NCB1\_TaxID=90311;  
 RN (1)

RP SEQUENCE FROM N.A.  
 RA Oriol R.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Candelier J.J., Martinez-Duncker I., Oriol R., Mollicone R.;  
 RT "Cloning expression and genomic organization of a new human alpha3-  
 RT fucosyltransferase (FUT10)."  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ535692; CAD59689.1;  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008417; F:fucosyltransferase activity; IEA.  
 DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.  
 DR InterPro; IPR001503; Glyco\_transf\_10;  
 DR Pfam; PF00852; Glyco\_transf\_10; I.  
 DR Transferrase; Glycosyltransferase.  
 KW SEQUENCE 439 AA; 51136 MW; 9A6B49B8E0C6E7EF CRC64;  
 SQ  
 Query Match 60.4%; Score 1586; DB 13; Length 439;  
 Best Local Similarity 71.8%; Pred. No. 1,4e-124;  
 Matches 293; Conservative 44; Mismatches 67; Indels 4; Gaps 2;  
 QY 34 EVWVELGKPERKEFKSSSLQDGHTEAAPTILNFKKGLTFRRKRWELDSYPTLW 93  
 DB 27 QVITELGNSENKAPAVSLHSGPLKPDENH---SOLKKNELYSFRTEPDTDHYPIILW 83  
 QY 94 WSLPLGTGRLGCCGADACFPITNRYLHNTKAPLYGIDFNIDSLPRKKAHDMAY 153  
 DB 84 WSLPLGTGSGCGGADVCFPIINKYQHSQMTAFPLFGYDFSIDSLPRKKAHDMAL 143  
 QY 154 FHESPPNNYKLPFKKVIITLFPNTATFSSHSLPTTQYLSIEVLKSLRYLVPLOSKN 213  
 DB 144 FHESPPNNYKLPFKKVIITLFPNTATFSSHSLPTTQYLSIEVLKSLRYLVPLOSKN 203  
 QY 214 LRRLAPLVYVQSDCPPSPDRDSYVRELMTYIEVDSYECIRNKDLPQOLKNPASMADG 273  
 DB 204 LRRLAPLVYVQSDCPPSPDRDSYVRELMTYIEVDSYECIRNKDLPQOLKNPASMADG 263  
 QY 274 FVRILAOYKFLAFENAVDDYITEKFWAPLKGVPVYQSPSTIDLPENKKAIIYSE 333  
 DB 264 FYKILAOYKFLAFENAVDDYITEKFWAPLKGVPVYQSPSTIDLPENKKAIIYSS 323  
 QY 334 FSHPRELASYIRRLSDRLYEAYEWKLGKISNORLLTLRERKMGVQDVNDYIDA 393  
 DB 324 FSHPRELASYIRRLSDRLYEAYEWKLGKISNORLLTLRERKMGVQDVNDYIDA 383  
 QY 394 FEQVCTKWANIRLOEKLPRKWEABDTHLSCEPVPVAFSPRLTP 441  
 DB 384 FEQVCTKWANIRLOEKLPRKWEABDTHLSCEPVPVAFSPRLTP 430  
 RESULT 7  
 Q81V16 PRELIMINARY; PRT; 357 AA.  
 AC Q81V16;  
 DT 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)  
 DE Putative alpha3-fucosyltransferase.  
 GN FUT10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Candelier J.J., Martinez-Duncker I., Oriol R., Mollicone R.;  
 RT "Cloning expression and genomic organization of two new human alpha3-  
 RT fucosyltransferases (FUT10 and FUT11)."  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ535839; CAD59712.1;  
 DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0008417; F:fucosyltransferase activity; IEA.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.  
 DR InterPro; IPR001503; Glyco\_transf\_10;  
 DR Pfam; PF00852; Glyco\_transf\_10; I.  
 DR Transferrase; Glycosyltransferase.  
 KW SEQUENCE 357 AA; 41347 MW; DD4F2FD085594D8 CRC64;  
 SQ  
 Query Match 57.1%; Score 1499; DB 4; Length 357;  
 Best Local Similarity 96.6%; Pred. No. 2.1e-117;  
 Matches 282; Conservative 1; Mismatches 3; Indels 6; Gaps 1;  
 QY 133 GIDFNIDSLPRKKAHDMAYFHESPPNNYKLPFKKVIITLFPNTATFSSHSLPTTQY 192  
 DB 64 GIDFNIDSLPRKKAHDMAYFHESPPNNYKLPFKKVIITLFPNTATFSSHSLPTTQY 123  
 QY 193 LESIVLKSRLYLVPLQSKNKLRLAPLVYVQSDCPPSPDRDSYVRELMTYIEVDSYGE 252  
 DB 124 LESIVLKSRLYLVPLQSKNKLRLAPLVYVQSDCPPSPDRDSYVRELMTYIEVDSYGE 183  
 QY 253 CLRNDLPQOLKNPASMADGFPYRIIAQYKFLAFENAVDDYITEKFWAPLKGVPVY 312  
 DB 184 CLRNDLPQOLKNPASMADGFPYRIIAQYKFLAFENAVDDYITEKFWAPLKGVPVY 243  
 QY 313 YGSPSITDMLPSNKSALIVSEFSPRELASYIRRLSDRLYEAYEWKLGKISNORLL 372  
 DB 244 YGSPSITDMLPSNKSALIVSEFSPRELASYIRRLSDRLYEAYEWKLGKISNORLL 303  
 QY 373 TALREKMGVQDVNDYIDAFECVCTKWANIRLOE-----GLPPKRW 418  
 DB 304 TALREKMGVQDVNDYIDAFECVCTKWANIRLOEYKSKVGIEPAGW 355  
 RESULT 8  
 Q8R247 PRELIMINARY; PRT; 264 AA.  
 AC Q8R247;  
 DT 01-JUN-2002 (TREMblrel. 21, Created)  
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC022579; AAH22579.1;  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008417; F:fucosyltransferase activity; IEA.  
 DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.  
 DR InterPro; IPR001503; Glyco\_transf\_10;  
 DR Pfam; PF00852; Glyco\_transf\_10; I.  
 KW Hypothetical protein.  
 FT NON TER  
 SQ SEQUENCE 264 AA; 30749 MW; GDE96A68DA25EB13 CRC64;  
 Query Match 42.5%; Score 1115.5; DB 11; Length 264;  
 Best Local Similarity 83.7%; Pred. No. 2.2e-85;  
 Matches 205; Conservative 20; Mismatches 17; Indels 3; Gaps 1;  
 QY 180 FSRSHSLPTTQYLSIEVLKSLRYLVPLQSKNKLRLAPLVYVQSDCPPSPDRDSYR 239  
 DB 1 FSRSHSLPTTQYLSIEVLKSLRYLVPLQSKNKLRLAPLVYVQSDCPPSPDRDSYR 60  
 QY 240 ELMTYIEVDSYECIRNKDLPQOLKNPASMADGFPYRIIAQYKFLAFENAVDDYITEK 299  
 DB 61 ELMTYIEVDSYECIRNKDLPQOLKNPASMADGFPYRIIAQYKFLAFENAVDDYITEK 120  
 QY 300 FWRPLKGVVPPVYGGSSITDMLPSNKSALIVSEFSPRELASYIRRLSDRLYEAYE 359

Db 121 FWRPLKGVPEVYGGSPFIADMLPSNRSAIVSEFHSHPRELAFIRLDYDGLYETYE 180  
 QY 360 WKLGKISNQLLTALRERKKGVDVNDVNTIAFEQMCVCTKWMANIRLOEKLPPKRE 419  
 Db 181 WKLGKISNQLLTALRERKKGVDVNDVNTIDFBCMCVRWMAUSRLQEQ--VSEWK 237  
 QY 420 AEDTH 424  
 Db 238 SGWTH 242

## RESULT 9

Q8AWC7 PRELIMINARY; PRT; 505 AA.  
 ID Q8AWC7  
 AC Q8AWC7  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, last sequence update)  
 DT 01-JUN-2003 (Tremblrel. 24, last annotation update)  
 GN Putative alpha3-fucosyltransferase.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NC NCBITaxid=9031;  
 RX NCBI\_Taxid=9031;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Candelier J.J., Martinez-Duncker I., Oriol R., Mollicone R.;  
 RT "Cloning expression and genomic organization of a new human alpha3-  
 fucosyltransferase (FUT10)."  
 RT Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AJ535752; CAD59736.1;  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008417; F:fucosyltransferase activity; IEA.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.  
 DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.  
 DR InterPro: IPR001503; Glyco. trans. 10.  
 DR Pfam: PF00852; Glyco. trans. 10.  
 DR Transferrase; Glycosyltransferase.  
 QX SEQUENCE 505 AA; 57516 MW; CA5102A0537B37F6 CRC64;

Query Match 34.2%; Score 899; DB 13; Length 505;  
 Best Local Similarity 44.6%; Pred. No. 7.5e-67;  
 Matches 182; Conservative 60; Mismatches 146; Indels 20; Gaps 7;

QY 89 PLTMTWS-----PLTGETGLGCGGACGFTTNRITLHHMTKALFLFGTDNITSLPL 143  
 Db 88 PLTMTWSGSLPFPHPGDTERT-DCPRGCLVTRSRRAARHRTKALIFGTDPRAYEARL 146  
 QY 144 PEKAHMDAVPEEHPKKNYKLPKPVITLFNTATFSRSHSLPLTQYESTSEVLSLR 203  
 Db 147 PLPHQTALPHEEPPMNNYLLSHPGIRLPNTYATFRRESDYPLTIQWLPAGYLRG-- 204  
 QY 204 YLVPLQSKKLRKR-LAPLVYVQSDCPSPDRDSYVELMTYIEVDSYGECLANKDLPOQ 262  
 Db 205 PALPLAEKDAWRRRRGYGVLYQSHCDVPSDRDYVELMKYIQVDSYGCILNRELPSH 264  
 QY 263 L---KNPMSMDADGYYRIIAQYKFLAFENAVCDQDYTEKFWPKLGVVYVYGGSPSTI 319  
 Db 265 RIRDTSTATEDPEPEVARIARKFHLAENAIQNDYMTKLMRPMHIGAVPVIRGSDYAR 324  
 QY 320 DWLPENKSAIVSEFHSHPRELASYIRRLSDDRLYEAVYEWKLGKISNQLLTALRERK 379  
 Db 325 DWMPNNLSITLLDDPDSQELAKYLDPLDKNGEYVYKLEYKNDG;KNGFLLESERRR 384  
 QY 380 WGVQVNDVNTIDAECVCTKWMANIRLOE-----KGLPPEKMAEATHTLSCEPFLV 432  
 Db 385 WGVNDVNTLPNYLNGECFICDRENNARYAEQEHKKSNGKTPAPSPHIAHQHNGCPMPTP 444  
 QY 433 FAFSPLRTPPLSLREMWISSFEQSKKAQALRWLVDRNCFSSQEFW 480  
 Db 445 GFGSYEDLPGBDSWKEMMLQDYWGSLDQGEALTMAMHNESHQGR-FW 491

## RESULT 10

Q81YE4 PRELIMINARY; PRT; 492 AA.  
 ID Q81YE4  
 AC Q81YE4  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, last sequence update)  
 DT 01-JUN-2003 (Tremblrel. 24, last annotation update)  
 DE Similar to putative alpha 1,3-fucosyl transferase.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NC NCBITaxid=9606;  
 RX NCBI\_Taxid=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Tissue=Testis;  
 RA Strausberg R.;  
 RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC036037; AAH36037.1;  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008417; F:fucosyltransferase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.  
 DR InterPro: IPR001503; Glyco. trans. 10.  
 DR Pfam: PF00852; Glyco. trans. 10; I.  
 DR Transferrase.  
 QX SEQUENCE 492 AA; 55800 MW; B6B3B5B0CB3B3C50 CRC64;

Query Match 33.2%; Score 872; DB 4; Length 492;  
 Best Local Similarity 42.6%; Pred. No. 1.3e-64;  
 Matches 178; Conservative 74; Mismatches 146; Indels 20; Gaps 9;

QY 79 RKKWEIDSYPIIMWSP-----LTGETGLGCGGACGFTTNRITLHHMTKALFLFG 133  
 Db 66 RRGEEEGDPLVILWSPGLPFPHPGDSERI-BCAGACVASHNRALDSRLLFLFG 124  
 QY 134 TDFNIDSLPEPKAHMDAVPEEHPKKNYKLPKPVITLFNTATFSRSHSLPLTQYL 193  
 Db 125 TDFASAPPLPRLAHQKALHRESPLNNFLSHGQIRLPNTSFSHSDVPLSLQWL 184  
 QY 194 ESIEVLKSLRYVPLQSKKLRKR-LAPLVYVQSDCPSPDRDSYVELMTYIEVDSYGE 252  
 Db 185 PGTAYLR--RPVPPMERAEWRRRGYAPLLYLQSHCDVPSDRDYVELMRHPIVDSYK 242  
 QY 253 CLRKNDLP-QOLKPAEMDADG--FYRIIAQYKFLAFENAVCDQDYTEKFWPKLGVV 309  
 Db 243 CLQRELPPTALQDTAATTEDEPLLAFLSKYFHLAENAIQNDYMTKLMRPMHIGAV 302  
 QY 310 PYYGSPSITDWLPENKSAIVSEFHSHPRELASYIRRLSDDRLYEAVYEWKLGKISNQ 369  
 Db 303 PYRGSPSVRRMNNHNSVILIDPFESPQKLAFFIDLDKNDEYKYLAYKQPGGITNQ 362  
 QY 370 RLTLALRERKKGVDVNDVNTIDAECVCTKWMANIRLOE-----KGLP--KRWAE 422  
 Db 363 FLDSLKRKRGVNDPLIPYLNQFECFVCDYELARIDAKAAAPGDSVPEPHIAP 422  
 QY 423 THLSCEPTEVAFPLRTPPLSLREMWISSFEQSKKAQALRWLVDRNCFSSQEFW 480  
 Db 423 SHMCPVTPPGFQGVVEIPEPDSWKEMMLQDYWGSLDQGEALTMAMHNHNT-EQTFW 479

## RESULT 11

Q8BHC9 PRELIMINARY; PRT; 489 AA.  
 ID Q8BHC9  
 AC Q8BHC9  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
 DE Hypothetical zinc finger (Putative alpha 1,3)  
 DE fucosyltransferase).  
 GN 311009G03RIK OR FUT11.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Brain, Ovary, and Uterus;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium.  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL; TISSUE=Kidney;  
RA Roos C.;  
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL; TISSUE=Kidney;  
RX MEDLINE=21671316; PubMed=11698403;  
RA Roos C., Kolmer M., Matilla P., Renkonen R.;  
RT "Composition of Drosophila melanogaster proteome involved in  
fucosylated glycan metabolism."  
RL J. Biol. Chem. 277:3168-3175 (2002).  
DR EMBL; AK030310; BAC26892.1; -  
DR EMBL; AJ542393; CAD62573.1; -  
DR PIR; P10559; P10677.  
DR MGI; MGI:1920318; 3110009G03Rik.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008417; F:fucosyltransferase activity; IEA.  
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. ; IEA.  
DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.  
DR InterPro; IPR001503; Glyco\_trans\_10.  
DR InterPro; IPR007087; ZnF\_C2H2.  
DR Pfam; PF00852; Glyco\_transf\_10; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
KW Hypothetical protein; Glycosyltransferase; Transferase.  
SQ SEQUENCE 489 AA; 55532 MW; 313E8E8A3A1345 CRC64;  
Query Match 32.8%; Score 862; DB 11; Length 489;  
Best Local Similarity 43.3%; Pred. No. 9,1e-64;  
Matches 178; Conservative 68; Mismatches 145; Indels 20; Gaps 9;  
QY 89 PIMLWSP-----LTGETGRIGCGACACFTTINRTYLLHHMTKAFLYFGTDFNIDSLPL 143  
DB 72 PVLWMSPLGFLPHFPDPSERI-QCAHGACVASRDRARADPRTRALLFYGTDFRADAPL 130  
QY 144 PRKAHDMVAVHEESPKNYKLFKPYVITLFTNTYTSRSHLLPLTQYLESTIVLSLR 203  
DB 131 PRLAHQSMALLHEESPLNFFLSHGPGIRLFTLTAFTSRHSDVPLQWLPGAAYLR--R 188  
QY 204 YLVLPLDSKNNKLRK-LAPLVVYQSDCDPSPDRSDSVREIMTYIEVDSYGECLRNKDLPO- 261  
DB 189 PAPPRERAEMWRKRGAPLVIYQSHCDVPSDRDRVRELMRYIEVDSYGECLQNRPEPTV 248  
QY 262 QLKNPASMDAG--FYRIIAQYKFLAENAVCDYITEKFWPLKGVVPVYVYGSPTIT 319  
DB 249 RLQDTATATTEDEPLMAFLSRYPKFLALENAICNDYVTEKLMRPMHIGAVPVYRGSPSVR 308  
QY 320 DMLPSNKSAIIVSEFSPRELASYIRRLSDSDRLLEAVVEWKLKGEISNORLLTALREK 379  
DB 309 DMKPNHSHVILLIDFESSPQKLAEFLDPLDKNDDEYKYLAVKQGGITNOFLDNLHRE 368  
QY 380 WGVQDVQNDYIDAEFCWVCTKWMANIRLQEK-----GLP-PRKWEADTHLSCEPPTV 432  
DB 369 WGVNDDEPLNLYNGECFVCDHEILALNAEKALHASSHGDIPVPPRIAQSSHMCPVPTP 428  
QY 433 FAFSPILRTPPLSLIREMWTSSFEQSKKAQALRWLVDRNQNFSSQDFWGLV 483  
DB -429 GFGKVEELIENDSWKEMVLQDIWQGLYQGEALTATINHNET-QQRKFWDYV 478

RESULT 12  
Q9CXS9  
ID Q9CXS9 PRELIMINARY; PRT; 489 AA.  
AC Q9CXS9;  
DT 01-JUN-2001 (Tremblrel, 17, Created)  
DT 01-JUN-2001 (Tremblrel, 17, Last sequence update)  
DT 01-JUN-2003 (Tremblrel, 24, Last annotation update)  
DE 3110009G03Rik protein.  
GN 3110009G03Rik.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryo;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Mikado I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombaerts P.,  
RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:688-690 (2001).  
DR EMBL; AK014029; BAB29123.1; -  
DR MGI; MGI:1920318; 3110009G03Rik.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008417; F:fucosyltransferase activity; IEA.  
DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.  
DR InterPro; IPR007087; ZnF\_C2H2.  
DR InterPro; IPR001503; Glyco\_trans\_10.  
DR Pfam; PF00852; Glyco\_transf\_10; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
SQ SEQUENCE 489 AA; 55789 MW; B4C75C586832791 CRC64;  
Query Match 32.8%; Score 861; DB 11; Length 489;  
Best Local Similarity 43.3%; Pred. No. 1.1e-63;  
Matches 178; Conservative 68; Mismatches 145; Indels 20; Gaps 9;  
QY 89 PIMLWSP-----LTGETGRIGCGACACFTTINRTYLLHHMTKAFLYFGTDFNIDSLPL 143  
DB 72 PVLWMSPLGFLPHFPDPSERI-QCAHGACVASRDRARADPRTRALLFYGTDFRADAPL 130  
QY 144 PRKAHDMVAVHEESPKNYKLFKPYVITLFTNTYTSRSHLLPLTQYLESTIVLSLR 203  
DB 131 PRLAHQSMALLHEESPLNFFLSHGPGIRLFTLTAFTSRHSDVPLQWLPGAAYLR--R 188  
QY 204 YLVLPLDSKNNKLRK-LAPLVVYQSDCDPSPDRSDSVREIMTYIEVDSYGECLRNKDLPO- 261  
DB 189 PAPPRERAEMWRKRGAPLVIYQSHCDVPSDRDRVRELMRYIEVDSYGECLQNRPEPTV 248  
QY 262 QLKNPASMDAG--FYRIIAQYKFLAENAVCDYITEKFWPLKGVVPVYVYGSPTIT 319  
DB 249 RLQDTATATTEDEPLMAFLSRYPKFLALENAICNDYVTEKLMRPMHIGAVPVYRGSPSVR 308  
QY 320 DMLPSNKSAIIVSEFSPRELASYIRRLSDSDRLLEAVVEWKLKGEISNORLLTALREK 379  
DB 309 DMKPNHSHVILLIDFESSPQKLAEFLDPLDKNDDEYKYLAVKQGGITNOFLDNLHRE 368

```

QY 380 WGVQVNDNDYIDAFECMVCCTKWANIRLOEK-----GLP-PKRWEDTHTLSCPEPTV 432
DB 369 WGVNDPMLPNYLNGBECFVCDHELAIRNAKAAHSSGDIPEVEPRIAOSSHNCCVPIIP 428
QY 423 FAFSPLRTPPLSSIREMWSISFEQSKKEAQAALRWLVDRNCFSSQEFWGLV 483
DB 429 GFGKVEIPEENDSWKEMWLODYWGGLYQGEALTMAMHNET-QQRKFWDYV 478

RESULT 13
Q8CG40 PRELIMINARY; PRT; 493 AA.
AC Q8CG40;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Putative alpha3-fucosyltransferase.
GN FUT11.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rattus norvegicus;
RA Candelier J.U., Martinez-Duncker I., Oriol R., Mollicone R.;
RT "Cloning, expression and genomic organization of a new human alpha3-
RT fucosyltransferase (FUT10).";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
EMBL, AJ535753; CAD59737.1;
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008417; F:fucosyltransferase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.
DR InterPro: IPR001503; Glyco trans 10.
DR Pfam: PF00652; Glyco trans 10; I.
DR Transferase; Glycosyltransferase.
KW SEQUENCE 493 AA; 55841 MW; 2993761C0860140B CRC64;
SQ

Query Match 32.4%; Score 851; DB 11; Length 493;
Best Local Similarity 43.1%; Pred. No. 7,7e-63;
Matches 177; Conservative 68; Mismatches 146; Indels 20; Gaps 9;

QY 89 PMLWSP----LTGTRGLGCGCAGD-CFTINTYTHHMTALFLYGTDPNIDSLP 143
DB 76 PVLWWSPOLPFPHPGDSERI-ECALGACVARSRRARADPRTALLFYGTDFRADAPL 134
QY 144 PKRAHNDMAVFHEESPKNKYKLFKPVITTLFNATFGRSHSLPLTQYJESIEVLKSLR 203
DB 135 PRLAHOSNALHESBPPLNLFSLHSGPGLRNLNTATFGRSHDYLLPQLMPLGAAVLR--R 192
QY 204 YLVPLQSKNKLKR-LAPLVYQSDCDPPSDSYVRELMTYIVDSYGCCLRNKDLPO- 261
DB 193 PAPPLAEERAEHRRRGVAPLLYQSHCDVPDRDRYVELMRYIVDSYGKCLQNRBPTV 252
QY 262 QLKNPASMDAG--FYRIIAQYKFLAFENAVCDYITEKFWRLKGVAVVYVYSSSIT 319
DB 253 RLQDTATATTEPELMAFLSRKYLFLAEKALCNDYTEKLRPMHIGAVPYVYSSSVR 312
QY 320 DMLPSNKAALVSEFSPRELASYIRLDSDDLRYEAVYEWKLXGBISNQLTALREK 379
DB 313 DMVNNHSAVILLDDPESQKLABFIDPLDKNDDEYWKYLAQKQGITNCLDNLTHRE 372
QY 380 WGVQVNDNDYIDAFECMVCCTKWANIRLOE-----GLP-PKRWEDTHTLSCPEPTV 432
DB 373 WGVNDPMLPNYLNGBECFVCDHELAIRNAKAAHSSGDIPEVEPRIAOSSHNCCVPIIP 432
QY 433 FAFSPLRTPPLSSIREMWSISFEQSKKEAQAALRWLVDRNCFSSQEFWGLV 483
DB 433 GFGKVEIPEENDSWKEMWLODYWGGLYQGEALTMAMHNET-QQRKFWDYV 482

RESULT 14

```

```

Q81928
ID Q81928 PRELIMINARY; PRT; 444 AA.
AC Q81928;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Alpha 1,3 fucosyltransferase.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Petit D., Picard F., Dupuy F., Germot A., Julien R., Maftah A.;
RT "Core a3- and a6-fucosyltransferases in Drosophila: characterization
RT and origin of diversity.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
EMBL, AY061932; AA131643.1;
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008417; F:fucosyltransferase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.
DR InterPro: IPR001503; Glyco trans 10.
DR Pfam: PF00652; Glyco trans 10; I.
DR Transferase; Glycosyltransferase.
KW SEQUENCE 444 AA; 51800 MW; 6D95748C85F2A76 CRC64;
SQ

Query Match 27.1%; Score 711; DB 5; Length 444;
Best Local Similarity 37.1%; Pred. No. 3,7e-51;
Matches 156; Conservative 80; Mismatches 154; Indels 30; Gaps 10;

QY 68 SFLKKEGTFNRKKKMLDYSPI-MLWMS-PLUGERGLGCGCAGD-CFTINTYTHHNM 125
DB 28 NIIVENIKFNS-----PVELWWSRDSMWNVDVQRCQIHICRLTKKS--RRPW 76
QY 126 TKALFVGCTDENDISLPLPKRAHNDMAVFHEESPKNKYKLFKPVITTLFNATFGRSH 185
DB 77 ARGVLFGSNKTKGDFLPRNHEGYWALLHESBRNRPFSNKEPLRHFFSTFSGYSN 136
QY 186 PLPTQYLESIEVLKSLRYLVPLQSKNKLKR-LAPLVYQSDCDPPSDSYVRELMTY 244
DB 137 LPLTMYLPSGEALTSXDYYTFPDGSKSGYRPSISVFPQSDCTWSGDEYVKELMKH 196
QY 245 IEVSYGECLEANKLPOLQKNP--ASMDAGFYRIIAQYKFLAFENAVCDYITEKFW 302
DB 197 LPIDSGSCLNRDLPESLQKQYINLYSPBLRLFLSEYKFMALIEVNAACPDYITEKFW 256
QY 303 PUKGVAVVYVYSSPSIDWLPSNKAALVSEFSPRELASYIRLDSDDLRYEAVYEWKL 362
DB 257 PLMGVIRIYIGSPFLIKDWEENKSAIFNDFQNPQALVEYLNKLADNKKLYNSYRQHKL 316
QY 363 --KGEISNQLITALEKMGVQVND--NYIDAFECMVCCTKWANIRLOEKGLPRKE 419
DB 317 NRRNPISNKLHLNLYRQYHIGDSSPGASLFEKEBCAVCYHYNIR-----NVK 367
QY 420 AADTLSCPEPTVFAFSPLRTPPLSSIREMWSISFEQSKKEAQAALRWLVDRNCFSSQEF 479
DB 368 ADLRHYNPLPBYAKMGCKIPQNVAD--WRAMEVGGQCAKLDFFPRRDIGFNDAEF 425

RESULT 15
Q9BSR3 PRELIMINARY; PRT; 143 AA.
ID Q9BSR3
AC Q9BSR3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;

```

RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Ovary;  
RA Strausberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC004884; AA04884.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 143 AA; 16640 MW; 40BAAC7220D81977 CRC64;

Query Match 21.0%; Score 552.5; DB 4; Length 143;  
Best Local Similarity 93.6%; Pred. No. 1.7e-38;  
Matches 102; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

|    |    |   |     |
|----|----|---|-----|
| Qy | 34 | EVNVEIGKFKERKFKSSLSODGHTKMEAPTHINSFLKKEGLTFNRRKRWELDSYPIMLW | 93  |
| Db | 27 | QNVVELGKFKERKFKSSLSODGHTKMEAPTHINSFLKKEGLTFNRRKRWELDSYPIMLW | 86  |
| Qy | 94 | WSPLTGEGTGRIGQCGADACFTINRTYLHHMTKAFLEPYG-TDFNIDSL           | 141 |
| Db | 87 | WSPLTGEGTGRIGQCGADACFTINRTYLHHMTKAFLEPYGQDFRLSPL            | 135 |

Search completed: September 9, 2004, 21:46:09  
Job time : 124 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 21:30:25 ; Search time 25 Seconds  
(without alignments)  
1012.244 Million cell updates/sec

Title: US-10-080-960-2  
Perfect score: 2625  
Sequence: 1 MKVTGPPQGVDSMQCFENDQ.....LVDRNQNFSSQEFWGLVFXD 486

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description       |
|------------|-------|-------------|--------|----|-------------------|
| 1          | 710.5 | 27.1        | 443    | 1  | Q9VLA1 drosophila |
| 2          | 300.5 | 11.4        | 503    | 1  | Q9VLA1 drosophila |
| 3          | 268.5 | 10.2        | 365    | 1  | Q9VLA1 drosophila |
| 4          | 263   | 10.0        | 359    | 1  | Q9VLA1 drosophila |
| 5          | 245.5 | 9.7         | 359    | 1  | Q9VLA1 drosophila |
| 6          | 244.5 | 9.3         | 374    | 1  | Q9VLA1 drosophila |
| 7          | 239.5 | 9.1         | 361    | 1  | Q9VLA1 drosophila |
| 8          | 234.5 | 8.9         | 372    | 1  | Q9VLA1 drosophila |
| 9          | 233.5 | 8.9         | 389    | 1  | Q9VLA1 drosophila |
| 10         | 230   | 8.8         | 401    | 1  | Q9VLA1 drosophila |
| 11         | 224   | 8.5         | 342    | 1  | Q9VLA1 drosophila |
| 12         | 219   | 8.3         | 405    | 1  | Q9VLA1 drosophila |
| 13         | 213   | 8.0         | 433    | 1  | Q9VLA1 drosophila |
| 14         | 209   | 8.0         | 501    | 1  | Q9VLA1 drosophila |
| 15         | 203   | 7.7         | 513    | 1  | Q9VLA1 drosophila |
| 16         | 199   | 7.6         | 433    | 1  | Q9VLA1 drosophila |
| 17         | 181.5 | 6.9         | 425    | 1  | Q9VLA1 drosophila |
| 18         | 121.5 | 4.6         | 721    | 1  | Q9VLA1 drosophila |
| 19         | 116   | 4.4         | 888    | 1  | Q9VLA1 drosophila |
| 20         | 110.5 | 4.2         | 767    | 1  | Q9VLA1 drosophila |
| 21         | 110   | 4.2         | 488    | 1  | Q9VLA1 drosophila |
| 22         | 106.5 | 4.1         | 3061   | 1  | Q9VLA1 drosophila |
| 23         | 106.5 | 4.1         | 3063   | 1  | Q9VLA1 drosophila |
| 24         | 106   | 4.0         | 655    | 1  | Q9VLA1 drosophila |
| 25         | 104   | 4.0         | 817    | 1  | Q9VLA1 drosophila |
| 26         | 104   | 4.0         | 993    | 1  | Q9VLA1 drosophila |
| 27         | 103   | 3.9         | 979    | 1  | Q9VLA1 drosophila |
| 28         | 103   | 3.9         | 2103   | 1  | Q9VLA1 drosophila |
| 29         | 103   | 3.9         | 616    | 1  | Q9VLA1 drosophila |
| 30         | 102.5 | 3.9         | 533    | 1  | Q9VLA1 drosophila |
| 31         | 100.5 | 3.8         | 533    | 1  | Q9VLA1 drosophila |
| 32         | 100.5 | 3.8         | 1411   | 1  | Q9VLA1 drosophila |
| 33         | 99.5  | 3.8         | 4196   | 1  | Q9VLA1 drosophila |

|    |      |     |      |   |                   |
|----|------|-----|------|---|-------------------|
| 34 | 99   | 3.8 | 747  | 1 | Q9VLA1 drosophila |
| 35 | 97.5 | 3.7 | 809  | 1 | Q9VLA1 drosophila |
| 36 | 97.5 | 3.7 | 1345 | 1 | Q9VLA1 drosophila |
| 37 | 97   | 3.7 | 704  | 1 | Q9VLA1 drosophila |
| 38 | 96.5 | 3.7 | 805  | 1 | Q9VLA1 drosophila |
| 39 | 95.5 | 3.6 | 486  | 1 | Q9VLA1 drosophila |
| 40 | 95.5 | 3.6 | 1318 | 1 | Q9VLA1 drosophila |
| 41 | 95   | 3.6 | 749  | 1 | Q9VLA1 drosophila |
| 42 | 95   | 3.6 | 1722 | 1 | Q9VLA1 drosophila |
| 43 | 95   | 3.6 | 2339 | 1 | Q9VLA1 drosophila |
| 44 | 94   | 3.6 | 550  | 1 | Q9VLA1 drosophila |
| 45 | 94   | 3.6 | 700  | 1 | Q9VLA1 drosophila |

## ALIGNMENTS

RESULT 1  
ID FUBT DROME STANDARD; PRT; 443 AA.  
AC Q9VLA1;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Alpha-(1,3)-fucosyltransferase B (EC 2.4.1.-) (Galactoside 3-L-fucosyltransferase).  
GN FUCB OR CG4435.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Canton-S;  
RX MEDLINE=21359431; PubMed=11382750;  
RA Fabini G., Freilinger A., Altman F., Wilson I.B.H.;  
RT "Identification of core alpha,3-fucosylated glycans and cloning of the requisite fucosyltransferase cDNA from Drosophila melanogaster: potential basis of the neutral anti-horse radish peroxidase epitope.";  
RL J. Biol. Chem. 276:28058-28067 (2001).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D., George R.A., Lewis S.E., Li P.W., Hoskins R.A., Galle R.F., Amanatides P.G., Scherer S.E., Li P.W., Ashburner M., Henderson S.N., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P., Burris K.C., Busam D.A., Butler H., Cadieu E., Cantor A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de la Batis R., Delcher A., Deng Z., Dwyer A.D., Dew I., Dietz S.M., Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durkin K.D., Evansgill C.C., Ferraz C., Ferriere S., Fleischmann W., Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Jorgensen C., Kallal M., Katsuma F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A., Kimmel B.E., Kodali C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasero P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X., Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Moberg C., Morris D., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb U.M., Palazzolo M., Piatman G.S., Pan S., Pollard V., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RESULT 2  
AC FUT2 PROME STANDARD; PRT; 503 AA.  
ID FUT2 PROME  
AC 09VU19; OSYSYS  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Glycoprotein 3-alpha-L-fucosyltransferase A (EC 2.4.1.214) (Core  
alpha-1,3)-fucosyltransferase.  
GN FUT2A OR CG6869.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephyridae; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;  
EN [1]  
RP SEQUENCE FROM N.A., AND FUNCTION.  
RC SPRAIN=Cancon-S;  
RC MEDLINE=21359431; PubMed=11382750;  
RA Fahnli G., Freilinger A., Altman F., Wilson I.B.H.;  
RA "Identification of core alpha1,3-fucosylated glycans and cloning of  
the requisite fucosyltransferase cDNA from Drosophila melanogaster";  
RT Potential basis of the neutral anti-horseradish peroxidase epitope";  
RL J. Biol. Chem. 276:128058-128067(2001).  
EN [2]  
RP SEQUENCE FROM N.A.  
RC SPRAIN=Berketley;  
RC MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celtniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milgram D.L.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch L., Beasley E.M.,  
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Berns P.V., Berman B.P., Bhandari D., Bolchakov S.,  
RA Borokova D., Botchan M.R., Bouck Y., Brokstein P., Brotier P.,  
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A.A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Dong L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evansglistata C.C., Ferrara C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman I.U., Hernandez J.R., Houck J.,  
RA Hostin D., Huston K.A., Howland J., Wei M.-H., Idegawa C.,  
RA Jaitani M., Kalushon F., Kaplen G.H., Ke Z., Kennison J.A., Keithum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lascko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Minklov G., Mishina N.V., Moberg C., Morris J., Mostrel A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclad J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Styricks R., Teator C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinscock G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye Y., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
RP REVISIONS.  
RP MEDLINE=22426069; PubMed=12537572; Matthews B.B., Campbell K.S.,  
RA Miste S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochman S.B.,



RA Smith C.D., Tupy J.L., Whitefield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celinker S.E., de Grey A.D.N.J., Dysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RT Lewis S.E.,  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review".  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN  
RP SEQUENCE FROM N.A.  
RX STRAIN=Berkeley; TISSUE=Embryo;  
RX MEDLINE=22426066; PubMed=12537569;  
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champagne M.,  
RA George R.A., Guan H., Krommiller B., Pacleeb J.M., Park S., Wan K.H.,  
RA Rubin G.M., Celinker S.E.,  
RA Drosophila full-length cDNA resource.";  
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
CC -1- FUNCTION: Catalyzes alpha-1,3 glycosidic linkages.  
CC -1- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-[N-acetyl-beta-D-  
CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-  
CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-  
CC glucosaminyl]asparagine = GDP + N(4)-[N-acetyl-beta-D-  
CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-  
CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1->3)]-  
CC -N-acetyl-beta-D-glucosaminyl]asparagine.  
CC -1- COFACTOR: Manganese.  
CC -1- PATHWAY: Glycosylation.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
CC form in trans cisternae of Golgi (By similarity).  
CC -1- SIMILARITY: Belongs to the glycosyltransferase family 10.  
CC -1- CAUTION: Ref.4 sequence differs from that shown due to a  
CC frameshift in position 112.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AJ030245; CAC41641.1; -  
CC EMBL: AE003532; AAF49657.2; -  
CC EMBL: AY071324; AAL48946.1; ALT\_FRAME.  
CC FLYBASE: FBgn0036485; FUCTA.  
CC DR GO: GO:0000138; C:Golgi trans cisterna, ISS.  
CC DR GO: GO:0018392; F:glycoprotein 3-alpha-L-fucosyltransferase a. . .; IDA.  
CC DR GO: GO:000486; P:protein amino acid glycosylation; IDA.  
CC DR Pfam: PF00852; Glyco\_transf\_10; 1.  
CC DR Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane;  
CC KW Signal-anchor; Golgi stack. CYTOPLASMIC (POTENTIAL).  
CC FT DOMAIN 1 10  
CC FT TRANSMEM 11 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
CC FT  
CC FT LUMENAL, CATALYTIC (POTENTIAL).  
CC FT CARBOHYD 29 503  
CC FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CONFLICT 53 53 H->Y (IN REF. 1).  
CC FT SEQUENCE 503 AA; 59159 MW; D9DA6078E6010CC CRC64;  
CC  
CC Query Match 11.4%; Score 30.5; DB 1; Length 503;  
CC Best Local Similarity 3.05%; Pred. No. 3.2e-16;  
CC Matches 100; Conservative 44; Mismatches 129; Indels 55; Gaps 15;  
CC  
CC QY 106 QCGAACPFTTIRTYLHMHMTKAFIPYGTDFNIDSLPRAAHNDMAVFEEBPKNNYKL 165  
CC DB 199 KCPVPTCELTARDA--STADMLYDHYIPTGIRPNSKQVMNLYLECP--YHT 252  
CC QY 166 FHKPVITTFNYATPSRSHSLPL--TTQYLES--IEVLKSLRYLVLDQSKNKLRLAP 220

DB 253 QNVKPADAINATAYRDSSTIVADYEKQYVTVQOQEDDIY-----SVNTRK--- 303  
QY 221 LVYVSDCDPPSDSDSVRELMTYIEVDSYGECLRNKDLPOQKNPASMADGFRYRII-A 279  
DB 304 VAMFVNSCGANRGLQYAHLEQKYEVDYIGACGNFR-----GSRPADKCFEILDN 355  
QY 280 QYKPLAFENAVCDYITTEKFW-RPLKLGVPYPYGS-PBITTMDPSNKSAIIVSEFSR 337  
DB 356 DYKFLAFENSNCDYITTEKFEFVALNRVLPTWGAARPDEYVSAARRRYIHVDFSSP 415  
QY 338 RELASYIRLDDRLTEAVYEWLKGESNQ-----RLTA-----LPRKMGVDVNG 387  
DB 416 KELEAYLRIDHDELINSYFKMGIGEPINTYWCVCATLNEQLRKFRM----- 468  
QY 368 DNYIDAFECM---VC-TRVWANTRLOE 410  
DB 469 --YTDLDWMRGPGVCTTRSRWRFKARK 494  
CC  
CC RESULT 3  
CC ID FUT3\_BOVIN STANDARD; PRT; 365 AA.  
CC AC Q1126;  
CC DT 01-OCT-1996 (Ref. 34, Created)  
CC DT 15-JUL-1998 (Ref. 36, Last sequence update)  
CC DT 10-OCT-2003 (Ref. 42, Last annotation update)  
CC DE Galactoside 3(4)-L-fucosyltransferase (EC 2.4.1.65) (Blood group Lewis  
CC DE alpha-4-fucosyltransferase) (Lewis FT) (Fucosyltransferase 3) (FUCT-  
CC DE ITI) (FUTB).  
CC GN FUT3.  
CC OS Bos taurus (Bovine).  
CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
CC OC Bovidae; Bovinae; Bos.  
CC NCBI\_TaxID=9913;  
CC RN  
CC RP SEQUENCE FROM N.A.  
CC RX MEDLINE=97236840; PubMed=9079712;  
CC RA Oulmouden A., Miernicki A., Petit J.-M., Costache M., Paldic M.M.,  
CC RA Mollicone R., Orlol R., Julien R.;  
CC RT "Molecular cloning and expression of a bovine alpha(1,3)-  
CC RT fucosyltransferase gene homologous to a putative ancestor gene of the  
CC RT human FUT3-FUT5-FUT6 cluster.";  
CC RL J. Biol. Chem. 272:8764-8773(1997).  
CC CC -1- FUNCTION: May catalyze alpha-1,3 and alpha-1,4 glycosidic linkages  
CC CC involved in the expression of sialyl Lewis X and Lewis X/SSA-1  
CC CC antigens. It may be involved in blood group Lewis determination.  
CC CC -1- CATALYTIC ACTIVITY: GDP-beta-L-fucose + beta-D-galactosyl-(1->3)-  
CC CC N-acetyl-D-glucosaminyl-R = GDP + beta-D-galactosyl-(1->3)-[alpha-  
CC CC L-fucosyl-(1->4)]-N-acetyl-beta-D-glucosaminyl-R.  
CC CC -1- PATHWAY: Glycosylation.  
CC CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
CC CC form in trans cisternae of Golgi.  
CC CC -1- TISSUE SPECIFICITY: LIVER, KIDNEY, LUNG AND BRAIN.  
CC CC -1- MISCELLANEOUS: Also acts on the corresponding 1,4-galactosyl  
CC CC derivative, forming 1,3-L-fucosyl links.  
CC CC -1- SIMILARITY: Belongs to the glycosyltransferase family 10.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X87810; CAA61079.1; -  
CC DR InterPro: IPR001503; Glyco\_transf\_10.  
CC DR Pfam: PF00852; Glyco\_transf\_10; 1.  
CC KW Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane;  
CC KW Signal-anchor; Golgi stack. CYTOPLASMIC (POTENTIAL).  
CC FT DOMAIN 1 14

FT TRANSMEM 15 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT DOMAIN 35 365 LUMENAL, CATALYTIC (POTENTIAL).  
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (PROBABLE).  
 SQ SEQUENCE 365 AA; 42654 MW, 18715A361B0025D3 CRC64;  
 Query Match 10.2%; Score 268.5; DB 1; Length 365;  
 Best Local Similarity 27.8%; Pred. No. 7e-14;  
 Matches 93; Conservative 52; Mismatches 126; Indels 63; Gaps 15;  
 QY 90 IMLWSPPLTGETGRIGQC-----GADACFTINRTYLLHHMTKAFLEFGTDFN---IDS 140  
 Db VLLMTWPF-NQPVALRSCBELWPGTADCOLTVNRS--EYPOADAVLVHREVSHRPQMQL 125  
 QY 141 LPLBRAHDMVAVHEESPKNVYKLFHKPVITLFNYTATFSHSHLPLTQYLSIEVLK 200  
 Db PPSPPRPGGQRWVFWSESPSCIKL--KQLDGYFULTWSYRDSDFMPYGMLEPMP-SQ 182  
 QY 201 SLRYVLPQSKNKLKRLAPLVYVQSDCDPPSDRDSYVRELMTYLEVDYGECLRNKDL 260  
 Db PVEVLNTLSAKTKL-----VAVVSNMNTDSIRVQYKLLKPHLDVYGRF--HTPLP 234  
 QY 261 QQLKRPASMDADGFYRIIQYKFLAFENAVCDYITTEKFR-PLKGVVYVYGSFSIT 319  
 Db 235 HAL-----MAKQLSQYKFLAFENSLHPDITETKMKNALQAMAVPVVLGPRVN 284  
 QY 320 --DMLPSNKSAIVSEFSPRELASYIRLSDSDDLAYEYEMKLGKISNQLTLALRE 377  
 Db 285 YEGFLRP-KAFIVHEFQSPKDLAQYLALDQVASYLNTYRWR-----ETLRP 332  
 QY 378 RKMGVQDVNDNYTDAFECMVCCTKRWANIRLOEK 411  
 Db 333 RSP-----SWALMFCACW--KLQOE 351  
 RESULT 4  
 FUTE\_HUMAN STANDARD; PRT; 359 AA.  
 ID FUTE\_HUMAN  
 AC P51993;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.65) (galactoside 3-L-fucosyltransferase) (Fucosyltransferase 6) (FUCT-VI).  
 GN FUTE OR FCT3A.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93077550; PubMed=1339443;  
 RA Weston B.W., Smith P.L., Kelly R.J., Lowe J.B.;  
 RT "Molecular cloning of a fourth member of a human alpha (1,3)fucosyltransferase gene family. Multiple homologous sequences that determine expression of the Lewis x, sialyl Lewis x, and sialyl Lewis x epitopes";  
 RT J Biol. Chem. 267:24575-24584(1992).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=93077550; PubMed=1339443;  
 RA Weston B.W., Smith P.L., Kelly R.J., Lowe J.B.;  
 RT "Molecular cloning of a fourth member of a human alpha (1,3)fucosyltransferase gene family. Multiple homologous sequences that determine expression of the Lewis x, sialyl Lewis x, and sialyl Lewis x epitopes";  
 RT J Biol. Chem. 267:24575-24584(1992).  
 RP [3]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=95378269; PubMed=7650030;  
 RA Cameron H.S., Szczepaniak D., Weston B.W.;  
 RT "Expression of human chromosome 19b alpha(1,3)-fucosyltransferase

RT genes in normal tissues. Alternative splicing, polyadenylation, and RT isoforms.";  
 RL J. Biol. Chem. 270:20112-20122(1995).  
 CC - FUNCTION: ENZYME INVOLVED IN THE BIOSYNTHESIS OF THE E-SELECTIN LIGAND. SIALYL-LEWIS X. CATALYZES THE TRANSFER OF FUCCOSE FROM GDP-BETA-FUCCOSE TO ALPHA-2,3 SIALYLATED SUBSTRATES.  
 CC - CATALYTIC ACTIVITY: GDP-beta-L-fucose + beta-D-galactosyl-(1->3)-N-acetyl-D-glucosaminyl-R = GDP + beta-D-galactosyl-(1->3)-[alpha-L-fucosyl-(1->4)]-N-acetyl-beta-D-glucosaminyl-R.  
 CC - PATHWAY: Glycosylation.  
 CC - SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound form in trans cisternae of Golgi.  
 CC - ALTERNATIVE PRODUCTS:  
 CC - Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=P51993-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=P51993-2; Sequence=VSP\_001780;  
 CC - TISSUE SPECIFICITY: KIDNEY, LIVER, COLON, SMALL INTESTINE, BLADDER, UTERUS AND SALIVARY GLAND.  
 CC - SIMILARITY: Belongs to the glycosyltransferase family 10.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL, M98825; AAA9922.1; -  
 CC EMBL, L01688; AAB03078.1; -  
 CC EMBL, U27333; AAC50192.1; -  
 CC EMBL, U27334; AAC50193.1; -  
 CC EMBL, U27335; AAC50194.1; -  
 CC EMBL, U27336; AAC50195.1; -  
 CC EMBL, U27331; AAC50190.1; -  
 CC EMBL, U27332; AAC50191.1; -  
 CC EMBL, U27337; AAC50196.1; -  
 CC EMBL, U27338; AAC50197.1; -  
 CC EMBL, U27339; AAC50198.1; -  
 CC EMBL, U27340; AAC50199.1; -  
 CC EMBL, U27341; AAC50200.1; -  
 CC EMBL, U27342; AAC50201.1; -  
 CC EMBL, U27343; AAC50202.1; -  
 CC EMBL, U27344; AAC50203.1; -  
 CC EMBL, U27345; AAC50204.1; -  
 CC EMBL, U27346; AAC50205.1; -  
 CC EMBL, U27347; AAC50206.1; -  
 CC EMBL, U27348; AAC50207.1; -  
 CC EMBL, U27349; AAC50208.1; -  
 CC EMBL, U27350; AAC50209.1; -  
 CC EMBL, U27351; AAC50210.1; -  
 CC EMBL, U27352; AAC50211.1; -  
 CC EMBL, U27353; AAC50212.1; -  
 CC EMBL, U27354; AAC50213.1; -  
 CC EMBL, U27355; AAC50214.1; -  
 CC EMBL, U27356; AAC50215.1; -  
 CC EMBL, U27357; AAC50216.1; -  
 CC EMBL, U27358; AAC50217.1; -  
 CC EMBL, U27359; AAC50218.1; -  
 CC EMBL, U27360; AAC50219.1; -  
 CC EMBL, U27361; AAC50220.1; -  
 CC EMBL, U27362; AAC50221.1; -  
 CC EMBL, U27363; AAC50222.1; -  
 CC EMBL, U27364; AAC50223.1; -  
 CC EMBL, U27365; AAC50224.1; -  
 CC EMBL, U27366; AAC50225.1; -  
 CC EMBL, U27367; AAC50226.1; -  
 CC EMBL, U27368; AAC50227.1; -  
 CC EMBL, U27369; AAC50228.1; -  
 CC EMBL, U27370; AAC50229.1; -  
 CC EMBL, U27371; AAC50230.1; -  
 CC EMBL, U27372; AAC50231.1; -  
 CC EMBL, U27373; AAC50232.1; -  
 CC EMBL, U27374; AAC50233.1; -  
 CC EMBL, U27375; AAC50234.1; -  
 CC EMBL, U27376; AAC50235.1; -  
 CC EMBL, U27377; AAC50236.1; -  
 CC EMBL, U27378; AAC50237.1; -  
 CC EMBL, U27379; AAC50238.1; -  
 CC EMBL, U27380; AAC50239.1; -  
 CC EMBL, U27381; AAC50240.1; -  
 CC EMBL, U27382; AAC50241.1; -  
 CC EMBL, U27383; AAC50242.1; -  
 CC EMBL, U27384; AAC50243.1; -  
 CC EMBL, U27385; AAC50244.1; -  
 CC EMBL, U27386; AAC50245.1; -  
 CC EMBL, U27387; AAC50246.1; -  
 CC EMBL, U27388; AAC50247.1; -  
 CC EMBL, U27389; AAC50248.1; -  
 CC EMBL, U27390; AAC50249.1; -  
 CC EMBL, U27391; AAC50250.1; -  
 CC EMBL, U27392; AAC50251.1; -  
 CC EMBL, U27393; AAC50252.1; -  
 CC EMBL, U27394; AAC50253.1; -  
 CC EMBL, U27395; AAC50254.1; -  
 CC EMBL, U27396; AAC50255.1; -  
 CC EMBL, U27397; AAC50256.1; -  
 CC EMBL, U27398; AAC50257.1; -  
 CC EMBL, U27399; AAC50258.1; -  
 CC EMBL, U27400; AAC50259.1; -  
 CC EMBL, U27401; AAC50260.1; -  
 CC EMBL, U27402; AAC50261.1; -  
 CC EMBL, U27403; AAC50262.1; -  
 CC EMBL, U27404; AAC50263.1; -  
 CC EMBL, U27405; AAC50264.1; -  
 CC EMBL, U27406; AAC50265.1; -  
 CC EMBL, U27407; AAC50266.1; -  
 CC EMBL, U27408; AAC50267.1; -  
 CC EMBL, U27409; AAC50268.1; -  
 CC EMBL, U27410; AAC50269.1; -  
 CC EMBL, U27411; AAC50270.1; -  
 CC EMBL, U27412; AAC50271.1; -  
 CC EMBL, U27413; AAC50272.1; -  
 CC EMBL, U27414; AAC50273.1; -  
 CC EMBL, U27415; AAC50274.1; -  
 CC EMBL, U27416; AAC50275.1; -  
 CC EMBL, U27417; AAC50276.1; -  
 CC EMBL, U27418; AAC50277.1; -  
 CC EMBL, U27419; AAC50278.1; -  
 CC EMBL, U27420; AAC50279.1; -  
 CC EMBL, U27421; AAC50280.1; -  
 CC EMBL, U27422; AAC50281.1; -  
 CC EMBL, U27423; AAC50282.1; -  
 CC EMBL, U27424; AAC50283.1; -  
 CC EMBL, U27425; AAC50284.1; -  
 CC EMBL, U27426; AAC50285.1; -  
 CC EMBL, U27427; AAC50286.1; -  
 CC EMBL, U27428; AAC50287.1; -  
 CC EMBL, U27429; AAC50288.1; -  
 CC EMBL, U27430; AAC50289.1; -  
 CC EMBL, U27431; AAC50290.1; -  
 CC EMBL, U27432; AAC50291.1; -  
 CC EMBL, U27433; AAC50292.1; -  
 CC EMBL, U27434; AAC50293.1; -  
 CC EMBL, U27435; AAC50294.1; -  
 CC EMBL, U27436; AAC50295.1; -  
 CC EMBL, U27437; AAC50296.1; -  
 CC EMBL, U27438; AAC50297.1; -  
 CC EMBL, U27439; AAC50298.1; -  
 CC EMBL, U27440; AAC50299.1; -  
 CC EMBL, U27441; AAC50300.1; -  
 CC EMBL, U27442; AAC50301.1; -  
 CC EMBL, U27443; AAC50302.1; -  
 CC EMBL, U27444; AAC50303.1; -  
 CC EMBL, U27445; AAC50304.1; -  
 CC EMBL, U27446; AAC50305.1; -  
 CC EMBL, U27447; AAC50306.1; -  
 CC EMBL, U27448; AAC50307.1; -  
 CC EMBL, U27449; AAC50308.1; -  
 CC EMBL, U27450; AAC50309.1; -  
 CC EMBL, U27451; AAC50310.1; -  
 CC EMBL, U27452; AAC50311.1; -  
 CC EMBL, U27453; AAC50312.1; -  
 CC EMBL, U27454; AAC50313.1; -  
 CC EMBL, U27455; AAC50314.1; -  
 CC EMBL, U27456; AAC50315.1; -  
 CC EMBL, U27457; AAC50316.1; -  
 CC EMBL, U27458; AAC50317.1; -  
 CC EMBL, U27459; AAC50318.1; -  
 CC EMBL, U27460; AAC50319.1; -  
 CC EMBL, U27461; AAC50320.1; -  
 CC EMBL, U27462; AAC50321.1; -  
 CC EMBL, U27463; AAC50322.1; -  
 CC EMBL, U27464; AAC50323.1; -  
 CC EMBL, U27465; AAC50324.1; -  
 CC EMBL, U27466; AAC50325.1; -  
 CC EMBL, U27467; AAC50326.1; -  
 CC EMBL, U27468; AAC50327.1; -  
 CC EMBL, U27469; AAC50328.1; -  
 CC EMBL, U27470; AAC50329.1; -  
 CC EMBL, U27471; AAC50330.1; -  
 CC EMBL, U27472; AAC50331.1; -  
 CC EMBL, U27473; AAC50332.1; -  
 CC EMBL, U27474; AAC50333.1; -  
 CC EMBL, U27475; AAC50334.1; -  
 CC EMBL, U27476; AAC50335.1; -  
 CC EMBL, U27477; AAC50336.1; -  
 CC EMBL, U27478; AAC50337.1; -  
 CC EMBL, U27479; AAC50338.1; -  
 CC EMBL, U27480; AAC50339.1; -  
 CC EMBL, U27481; AAC50340.1; -  
 CC EMBL, U27482; AAC50341.1; -  
 CC EMBL, U27483; AAC50342.1; -  
 CC EMBL, U27484; AAC50343.1; -  
 CC EMBL, U27485; AAC50344.1; -  
 CC EMBL, U27486; AAC50345.1; -  
 CC EMBL, U27487; AAC50346.1; -  
 CC EMBL, U27488; AAC50347.1; -  
 CC EMBL, U27489; AAC50348.1; -  
 CC EMBL, U27490; AAC50349.1; -  
 CC EMBL, U27491; AAC50350.1; -  
 CC EMBL, U27492; AAC50351.1; -  
 CC EMBL, U27493; AAC50352.1; -  
 CC EMBL, U27494; AAC50353.1; -  
 CC EMBL, U27495; AAC50354.1; -  
 CC EMBL, U27496; AAC50355.1; -  
 CC EMBL, U27497; AAC50356.1; -  
 CC EMBL, U27498; AAC50357.1; -  
 CC EMBL, U27499; AAC50358.1; -  
 CC EMBL, U27500; AAC50359.1; -  
 CC EMBL, U27501; AAC50360.1; -  
 CC EMBL, U27502; AAC50361.1; -  
 CC EMBL, U27503; AAC50362.1; -  
 CC EMBL, U27504; AAC50363.1; -  
 CC EMBL, U27505; AAC50364.1; -  
 CC EMBL, U27506; AAC50365.1; -  
 CC EMBL, U27507; AAC50366.1; -  
 CC EMBL, U27508; AAC50367.1; -  
 CC EMBL, U27509; AAC50368.1; -  
 CC EMBL, U27510; AAC50369.1; -  
 CC EMBL, U27511; AAC50370.1; -  
 CC EMBL, U27512; AAC50371.1; -  
 CC EMBL, U27513; AAC50372.1; -  
 CC EMBL, U27514; AAC50373.1; -  
 CC EMBL, U27515; AAC50374.1; -  
 CC EMBL, U27516; AAC50375.1; -  
 CC EMBL, U27517; AAC50376.1; -  
 CC EMBL, U27518; AAC50377.1; -  
 CC EMBL, U27519; AAC50378.1; -  
 CC EMBL, U27520; AAC50379.1; -  
 CC EMBL, U27521; AAC50380.1; -  
 CC EMBL, U27522; AAC50381.1; -  
 CC EMBL, U27523; AAC50382.1; -  
 CC EMBL, U27524; AAC50383.1; -  
 CC EMBL, U27525; AAC50384.1; -  
 CC EMBL, U27526; AAC50385.1; -  
 CC EMBL, U27527; AAC50386.1; -  
 CC EMBL, U27528; AAC50387.1; -  
 CC EMBL, U27529; AAC50388.1; -  
 CC EMBL, U27530; AAC50389.1; -  
 CC EMBL, U27531; AAC50390.1; -  
 CC EMBL, U27532; AAC50391.1; -  
 CC EMBL, U27533; AAC50392.1; -  
 CC EMBL, U27534; AAC50393.1; -  
 CC EMBL, U27535; AAC50394.1; -  
 CC EMBL, U27536; AAC50395.1; -  
 CC EMBL, U27537; AAC50396.1; -  
 CC EMBL, U27538; AAC50397.1; -  
 CC EMBL, U27539; AAC50398.1; -  
 CC EMBL, U27540; AAC50399.1; -  
 CC EMBL, U27541; AAC50400.1; -  
 CC EMBL, U27542; AAC50401.1; -  
 CC EMBL, U27543; AAC50402.1; -  
 CC EMBL, U27544; AAC50403.1; -  
 CC EMBL, U27545; AAC50404.1; -  
 CC EMBL, U27546; AAC50405.1; -  
 CC EMBL, U27547; AAC50406.1; -  
 CC EMBL, U27548; AAC50407.1; -  
 CC EMBL, U27549; AAC50408.1; -  
 CC EMBL, U27550; AAC50409.1; -  
 CC EMBL, U27551; AAC50410.1; -  
 CC EMBL, U27552; AAC50411.1; -  
 CC EMBL, U27553; AAC50412.1; -  
 CC EMBL, U27554; AAC50413.1; -  
 CC EMBL, U27555; AAC50414.1; -  
 CC EMBL, U27556; AAC50415.1; -  
 CC EMBL, U27557; AAC50416.1; -  
 CC EMBL, U27558; AAC50417.1; -  
 CC EMBL, U27559; AAC50418.1; -  
 CC EMBL, U27560; AAC50419.1; -  
 CC EMBL, U27561; AAC50420.1; -  
 CC EMBL, U27562; AAC50421.1; -  
 CC EMBL, U27563; AAC50422.1; -  
 CC EMBL, U27564; AAC50423.1; -  
 CC EMBL, U27565; AAC50424.1; -  
 CC EMBL, U27566; AAC50425.1; -  
 CC EMBL, U27567; AAC50426.1; -  
 CC EMBL, U27568; AAC50427.1; -  
 CC EMBL, U27569; AAC50428.1; -  
 CC EMBL, U27570; AAC50429.1; -  
 CC EMBL, U27571; AAC50430.1; -  
 CC EMBL, U27572; AAC50431.1; -  
 CC EMBL, U27573; AAC50432.1; -  
 CC EMBL, U27574; AAC50433.1; -  
 CC EMBL, U27575; AAC50434.1; -  
 CC EMBL, U27576; AAC50435.1; -  
 CC EMBL, U27577; AAC50436.1; -  
 CC EMBL, U27578; AAC50437.1; -  
 CC EMBL, U27579; AAC50438.1; -  
 CC EMBL, U27580; AAC50439.1; -  
 CC EMBL, U27581; AAC50440.1; -  
 CC EMBL, U27582; AAC50441.1; -  
 CC EMBL, U27583; AAC50442.1; -  
 CC EMBL, U27584; AAC50443.1; -  
 CC EMBL, U27585; AAC50444.1; -  
 CC EMBL, U27586; AAC50445.1; -  
 CC EMBL, U27587; AAC50446.1; -  
 CC EMBL, U27588; AAC50447.1; -  
 CC EMBL, U27589; AAC50448.1; -  
 CC EMBL, U27590; AAC50449.1; -  
 CC EMBL, U27591; AAC50450.1; -  
 CC EMBL, U27592; AAC50451.1; -  
 CC EMBL, U27593; AAC50452.1; -  
 CC EMBL, U27594; AAC50453.1; -  
 CC EMBL, U27595; AAC50454.1; -  
 CC EMBL, U27596; AAC50455.1; -  
 CC EMBL, U27597; AAC50456.1; -  
 CC EMBL, U27598; AAC50457.1; -  
 CC EMBL, U27599; AAC50458.1; -  
 CC EMBL, U27600; AAC50459.1; -  
 CC EMBL, U27601; AAC50460.1; -  
 CC EMBL, U27602; AAC50461.1; -  
 CC EMBL, U27603; AAC50462.1; -  
 CC EMBL, U27604; AAC50463.1; -  
 CC EMBL, U27605; AAC50464.1; -  
 CC EMBL, U27606; AAC50465.1; -  
 CC EMBL, U27607; AAC50466.1; -  
 CC EMBL, U27608; AAC50467.1; -  
 CC EMBL, U27609; AAC50468.1; -  
 CC EMBL, U27610; AAC50469.1; -  
 CC EMBL, U27611; AAC50470.1; -  
 CC EMBL, U27612; AAC50471.1; -  
 CC EMBL, U27613; AAC50472.1; -  
 CC EMBL, U27614; AAC50473.1; -  
 CC EMBL, U27615; AAC50474.1; -  
 CC EMBL, U27616; AAC50475.1; -  
 CC EMBL, U27617; AAC50476.1; -  
 CC EMBL, U27618; AAC50477.1; -  
 CC EMBL, U27619; AAC50478.1; -  
 CC EMBL, U27620; AAC50479.1; -  
 CC EMBL, U27621; AAC50480.1; -  
 CC EMBL, U27622; AAC50481.1; -  
 CC EMBL, U27623; AAC50482.1; -  
 CC EMBL, U27624; AAC50483.1; -  
 CC EMBL, U27625; AAC50484.1; -  
 CC EMBL, U27626; AAC50485.1; -  
 CC EMBL, U27627; AAC50486.1; -  
 CC EMBL, U27628; AAC50487.1; -  
 CC EMBL, U27629; AAC50488.1; -  
 CC EMBL, U27630; AAC50489.1; -  
 CC EMBL, U27631; AAC50490.1; -  
 CC EMBL, U27632; AAC50491.1; -  
 CC EMBL, U27633; AAC50492.1; -  
 CC EMBL, U27634; AAC50493.1; -  
 CC EMBL, U27635; AAC50494.1; -  
 CC EMBL, U27636; AAC50495.1; -  
 CC EMBL, U27637; AAC50496.1; -  
 CC EMBL, U27638; AAC50497.1; -  
 CC EMBL, U27639; AAC50498.1; -  
 CC EMBL, U27640; AAC50499.1; -  
 CC EMBL, U27641; AAC50500.1; -  
 CC EMBL, U27642; AAC50501.1; -  
 CC EMBL, U27643; AAC50502.1; -  
 CC EMBL, U27644; AAC50503.1; -  
 CC EMBL, U27645; AAC50504.1; -  
 CC EMBL, U27646; AAC50505.1; -  
 CC EMBL, U27647; AAC50506.1; -  
 CC EMBL, U27648; AAC50507.1; -  
 CC EMBL, U27649; AAC50508.1; -  
 CC EMBL, U27650; AAC50509.1; -  
 CC EMBL, U27651; AAC50510.1; -  
 CC EMBL, U27652; AAC50511.1; -  
 CC EMBL, U27653; AAC50512.1; -  
 CC EMBL, U27654; AAC50513.1; -  
 CC EMBL, U27655; AAC50514.1; -  
 CC EMBL, U27656; AAC50515.1; -  
 CC EMBL, U27657; AAC50516.1; -  
 CC EMBL, U27658; AAC50517.1; -  
 CC EMBL, U27659; AAC50518.1; -  
 CC EMBL, U27660; AAC50519.1; -  
 CC EMBL, U27661; AAC50520.1; -  
 CC EMBL, U27662; AAC50521.1; -  
 CC EMBL, U27663; AAC50522.1; -  
 CC EMBL, U27664; AAC50523.1; -  
 CC EMBL, U27665; AAC50524.1; -  
 CC EMBL, U27666; AAC50525.1; -  
 CC EMBL, U27667; AAC50526.1; -  
 CC EMBL, U27668; AAC50527.1; -  
 CC EMBL, U27669; AAC50528.1; -  
 CC EMBL, U27670; AAC50529.1; -  
 CC EMBL, U27671; AAC50530.1; -  
 CC EMBL, U27672; AAC50531.1; -  
 CC EMBL, U27673; AAC50532.1; -  
 CC EMBL, U27674; AAC50533.1; -  
 CC EMBL, U27675; AAC50534.1; -  
 CC EMBL, U27676; AAC50535.1; -  
 CC EMBL, U27677; AAC50536.1; -  
 CC EMBL, U27678; AAC50537.1; -  
 CC EMBL, U27679; AAC50538.1; -  
 CC EMBL, U27680; AAC50539.1; -  
 CC EMBL, U27681; AAC50540.1; -  
 CC EMBL, U27682; AAC50541.1; -  
 CC EMBL, U27683; AAC50542.1; -  
 CC EMBL, U27684; AAC50543.1; -  
 CC EMBL, U27685; AAC50544.1; -  
 CC EMBL, U27686; AAC50545.1; -  
 CC EMBL, U27687; AAC50546.1; -  
 CC EMBL, U27688; AAC50547.1; -  
 CC EMBL, U27689; AAC50548.1; -  
 CC EMBL, U27690; AAC50549.1; -  
 CC EMBL, U27691; AAC50550.1; -  
 CC EMBL, U27692; AAC50551.1; -  
 CC EMBL, U27693; AAC50552.1; -  
 CC EMBL, U27694; AAC50553.1; -  
 CC EMBL, U27695; AAC50554.1; -  
 CC EMBL, U27696; AAC50555.1; -  
 CC EMBL, U27697; AAC50556.1; -  
 CC EMBL, U27698; AAC50557.1; -  
 CC EMBL, U27699; AAC50558.1; -  
 CC EMBL, U27700; AAC50559.1; -  
 CC EMBL, U27701; AAC50560.1; -  
 CC EMBL, U27702; AAC50561.1; -  
 CC EMBL, U27703; AAC50562.1; -  
 CC EMBL, U27704; AAC50563.1;

```

Db      159 RSDSDIFPPYGMLEWWSGQPAHPPLNLSAKTEL-----VAMVSNMGPSARVRYQSL 212
QY      242 MTYIEVDSYGECLRNKDLPOQLKNPASMDAGFRIIAQYKFIAPENAVCDYITEKFM 301
Db      213 QAHKLVQVYGR--SKKPLPQ-----GTMETLSRYKFIAPFNSLHPDIYIEKLM 260
QY      302 R-PLKGVVPPVYYSPTDW---LPSNKSAILVSEFSHPRELAAYIRRLSDSDRLYEAV 357
Db      261 RNALAEAMAVPVLG-PSRSNYERFLPPD-AFIHVDDFQSPKDLARYLQELDKDHAYLSV 318
QY      358 VEWKLGKGEISNORLTLALRER--KMGVDVQNDNYIDAFCQVCTKWANIRLQEGKL 413
Db      319 FRWR-----ETLRPRFSKWL-----AF-CKACKWLQESRYQTGTGI 354

RESULT 5
FUT6_PANTR
ID_FUT6_PANTR STANDARD; PRT: 359 AA.
AC PS6434;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.65) (Galactoside 3-L-fucosyltransferase) (Fucosyltransferase 6) (FUCT-VI).
GN FUT6.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
NCBI_TaxID=9598;
RP SEQUENCE FROM N.A., AND VARIANTS LEU-124; GLU-172 AND ALA-192.
RX MEDLINE=98037800; PubMed=9368041;
RA Costache M., Apoli P.-A., Calileau A., Elmgren A., Larson G., Henry S., Blanche A., Iordachescu D., Ortol R., Mollicone R., "Evolution of fucosyltransferase genes in vertebrates." J. Biol. Chem. 272:29721-29728(1997).
RL J. Biol. Chem. 272:29721-29728(1997).
CC -1- FUNCTION: ENZYME INVOLVED IN THE BIOSYNTHESIS OF THE E-SELECTIN LIGAND. STALYI-LEWIS X. CATALYZES THE TRANSFER OF FUCOSE FROM GDP-BETA-FUCOSE TO ALPHA-2,3 SATURATED SUBSTRATES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: GDP-beta-L-fucose + beta-D-galactosyl-(1->3)-N-acetyl-D-glucosaminyl-R = GDP + beta-D-galactosyl-(1->3)-L-fucosyl-(1->4)-N-acetyl-beta-D-glucosaminyl-R.
CC -1- PATHWAY: Glycosylation.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound form in trans cisternae of Golgi (By similarity).
CC -1- POLYMORPHISM: There are two alleles A and B. Allele A has Pro-124, Gln-172 and Val-192. Allele B has Leu-124, Glu-172 and Ala-192.
CC -1- SIMILARITY: Belongs to the glycosyltransferase family 10.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL: Y14035; CAA74362.1;
DR InterPro: IPR001503; Glyco_transf_10.
DR Pfam: PF00852; Glyco_transf_10; I.
KM Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal-anchor; Golgi stack; Polymorphism.
KW Signal-anchor; Golgi stack; Polymorphism.
FT TRANSMEM 1 14
FT DOMAIN 1 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT CARBOHYD 35 359 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 46 46 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 184 184 N-LINKED (GLCNAc...) (POTENTIAL).
FT VARIANT 124 124 P->L (in allele B).

```

```

FT VARIANT 172 172 O->E (in allele B).
FT VARIANT 192 192 V->A (in allele B).
SQ SEQUENCE 359 AA; 41892 MW; 207C3DC29B2BFD3 CRC64;

Query Match 9.7%; Score 254; DB 1; Length 359;
Best Local Similarity 26.4%; Pred. No. 9.7e-13;
Matches 96; Conservative 45; Mismatches 118; Indels 104; Gaps 19;

QY 87 SYPLMLWSPITGCTGLGQC-----GADACFTIR-----TYLHHMTKALFLFG 133
Db 60 SIPPLIMTWPFNPFIALPRCSEWVGTADCNITARKYVQDAVIYHRE--VMN 115
QY 134 TDFNIDSLP-LPRKAHDMAVAFHEESPKNYKLFHKPVITFLFYATFSGH-----SHLP 187
Db 116 PS---AQLPSRPGQGRNITFMSIESFN-----CRHLEALDGYFN 153
QY 188 LFTQLESIEVLSKRLVLP-----LQSKNKRRLAPLYVQSDCDPSPDRDS 236
Db 154 LTMSYRSDSDIFPPYGMLEWWSGQPAHPPLNLSAKTEL-----VAMVSNMGPSARV 207
QY 237 YVREIMTYIEVDSYGECLRNKDLPOQLKNPASMDAGFRIIAQYKFIAPENAVCDYI 296
Db 208 YVQSLQAHKLVQVYGR--SKKPLPQ-----GTMETLSRYKFIAPFNSLHPDIY 255
QY 297 TEKFWR-PLKGVVPPVYYSPTDW---LPSNKSAILVSEFSHPRELAAYIRRLSDSDR 352
Db 256 TEXLMRNALAEAMAVPVLG-PSRSNYERFLPPD-AFIHVDDFQSPKDLARYLQELDKDH 313
QY 333 LYEAWEKLGKGEISNORLTLALRER--KMGVDVQNDNYIDAFCQVCTKWANIRLQEG 410
Db 314 RYLSYFRWR-----ETLRPRFSKWL-----AF-CKACKWLQESRYQTGTGI 351
QY 411 KGL 413
Db 352 RSI 354

RESULT 6
FUT5_HUMAN
ID_FUT5_HUMAN STANDARD; PRT: 374 AA.
AC Q11128;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.65) (Galactoside 3-L-fucosyltransferase) (Fucosyltransferase 5) (FUCT-V).
GN FUT5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RX TISSUE=peripheral blood leukocytes;
RX MEDLINE=92156161; PubMed=1740457;
RA Weston B.W., Nait R.P., Larsen R.D., Lowe J.B., "Isolation of a novel human alpha (1,3)fucosyltransferase gene and molecular comparison to the human Lewis blood group alpha (1,3/1,4)fucosyltransferase gene. Syntenic, homologous, nonallelic genes encoding enzymes with distinct acceptor substrate specificities." J. Biol. Chem. 267:4152-4160(1992).
RL J. Biol. Chem. 267:4152-4160(1992).
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=colon, kidney, and liver;
RX MEDLINE=95378269; PubMed=7650030;
RA Cameron H.S., Szepeanlak D., Weston B.W., "Expression of human chromosome 19p alpha(1,3)-fucosyltransferase gene in normal tissues. Alternative splicing, polyadenylation, and isoforms." J. Biol. Chem. 270:20112-20122(1995).
RL J. Biol. Chem. 270:20112-20122(1995).
CC -1- FUNCTION: May catalyze alpha-1,3 glycosidic linkages involved in the expression of VIM-2, Lewis X/SSA-1 and sialyl Lewis X

```

```
CC      -1- CATALYTIC ACTIVITY: GDP-beta-L-fucose + beta-D-galactosyl-(1->3) -
CC      N-acetyl-D-glucosaminyl-R = GDP + beta-D-galactosyl-(1->3) - [alpha-
CC      L-fucosyl-(1->4)]-N-acetyl-beta-D-glucosaminyl-R.
CC      -1- PATHWAY: Glycosylation.
CC      -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC      form in trans cisternae of Golgi.
CC      -1- TISSUE SPECIFICITY: LIVER, COLON AND TESTIS AND TRACE AMOUNTS IN
CC      -1- SIMILARITY: Belongs to the glycosyltransferase family 10.
CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M81485; AAA98117.1; -
DR      EMBL; U27329; AAC50188.1; -
DR      EMBL; U27330; AAC50189.1; -
DR      PIR; A42270; A42270.
DR      Genew; HGNC:4016; FUT5.
DR      MIM; 136835; -
DR      GO; GO:0008417; F-fucosyltransferase activity; TAS.
DR      GO; GO:0005975; Polysaccharide metabolism; TAS.
DR      InterPro; IPR001503; Glyco_transf_10.
DR      Pfam; Pf00852; Glyco_transf_10; 1.
CC      KMW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
CC      Signal-anchor; Golgi stack.
CC      KMW DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
CC      TRANSMEM 16 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC      (POTENTIAL).
CC      FT DOMAIN 35 374 LUMENAL, CATALYTIC (POTENTIAL).
CC      FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      SQ SEQUENCE 374 AA; 43008 MW; B825281521B57939 CRC64;
Query Match 9.4%; Score 245.5; DB 1; Length 374;
Best Local Similarity 26.6%; Pred. No. 4.9e-12;
Matches 94; Conservative 42; Mismatches 117; Indels 101; Gaps 19;
QY 90 IMLMWSPPLTGETGRIGOC-----GADACFTINR-----TYLHHHTKAFLEYGTDN 137
DB 78 ILMTWPFNTFVA-LPRCSEMYPGAADCNITLADSSVYQADAVIYHMD--IMYPSAN 133
QY 138 IDSLPLPKAHMDVAVHESPKNNYKLFHKPVITLFTNYATFSRH-----SHLPLTQY 192
DB 134 LP--PPTRPQGGQWIMFWSMSPSN-----CRHLALDGFNLTMSY 172
QY 193 LESIEVLKSLRIVP-----LOSKNLKRRLAPLVYVQSDCDPSDRSYREL 241
DB 173 RSDSDIFTPYGMWLEPMSGQPAHPPLNLAKTEL-----VAMAVSNMKPDSARVRYQSL 226
QY 242 MTYIEVDYSGECLRNKDLPOOLKNPASMADGFRYIIAOKFIILAFENAVCDYITEKFW 301
DB 227 QAHLLKVDYGR--SHKPLPK-----GTMMETLSRKYFLAFENSLHPDITTEKLM 274
QY 302 R-PLKGVVPVYVGSITDM--LPSNKSALIVSESHPRELASIYRRLDSDRLYEAY 357
DB 275 RNALKAAVAVPVVLG-PSRSNYERFLPPD-AFIHVDDFQSPKDLARYIOELDKHARLYSY 332
QY 358 VEWKLKGEISNORLLTALRE--KWGVQDVNDNDYIDAFECMCCTKWANIRLQ 409
DB 333 FRWR-----ETLPRPSFWAL-----AF-CKAKCKLQGESRYQ 364
RESULT 7
FUT5_PANTR
ID FUT5_PANTR STANDARD; PRT; 374 AA.
```

```
AC P56433;
DT 15-JUL-1998 (Rel. 36, Created)
DR 15-JUL-1998 (Rel. 36, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.65) (Galactoside 3-L-
DE fucosyltransferase) (Fucosyltransferase 5) (FUCT-V).
GN FUT5.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98037800; PubMed=9369041;
RX Costache M., Apoll P.-A., Gaillean A., Elmgren A., Larson G.,
RX Henry S., Blancher A., Iordachescu D., Oriol R., Mollicone R.;
RX "Evolution of fucosyltransferase genes in vertebrates.";
RX J. Biol. Chem. 272:29721-29728 (1997).
CC -1- FUNCTION: May catalyze alpha-1,3 glycosidic linkages involved in
CC the expression of VM-2, Lewis X/SSA-1 and sialyl Lewis X
CC antigens (By similarity).
CC -1- CATALYTIC ACTIVITY: GDP-beta-L-fucose + beta-D-galactosyl-(1->3) -
CC N-acetyl-D-glucosaminyl-R = GDP + beta-D-galactosyl-(1->3) - [alpha-
CC L-fucosyl-(1->4)]-N-acetyl-beta-D-glucosaminyl-R.
CC -1- PATHWAY: Glycosylation.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi.
CC -1- SIMILARITY: Belongs to the glycosyltransferase family 10.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Y14034; CAAT74361.1; -
DR      InterPro; IPR001503; Glyco_transf_10.
DR      Pfam; PF00852; Glyco_transf_10; 1.
CC      KMW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
CC      Signal-anchor; Golgi stack.
CC      KMW DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
CC      TRANSMEM 16 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC      (POTENTIAL).
CC      FT DOMAIN 35 374 LUMENAL, CATALYTIC (POTENTIAL).
CC      FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      SQ SEQUENCE 374 AA; 43034 MW; 3F35DEB8A8FA282 CRC64;
Query Match 9.3%; Score 244.5; DB 1; Length 374;
Best Local Similarity 26.3%; Pred. No. 5.0e-12;
Matches 93; Conservative 40; Mismatches 120; Indels 101; Gaps 18;
QY 90 IMLMWSPPLTGETGRIGOC-----GADACFTINR-----TYLHHHTKAFLEYGTDN 137
DB 78 ILMTWPFNTFVA-LPRCSEMYPGAADCNITLADSSVYQADAVIYHMD--IMYPSAN 133
QY 138 IDSLPLPKAHMDVAVHESPKNNYKLFHKPVITLFTNYATFSRH-----SHLPLTQY 192
DB 134 LP--PPTRPQGGQWIMFWSMSPSN-----CRHLALDGFNLTMSY 172
QY 193 LESIEVLKSLRIVP-----LOSKNLKRRLAPLVYVQSDCDPSDRSYREL 241
DB 173 RSDSDIFTPYGMWLEPMSGQPAHPPLNLAKTEL-----VAMAVSNMKPDSARVRYQSL 226
QY 242 MTYIEVDYSGECLRNKDLPOOLKNPASMADGFRYIIAOKFIILAFENAVCDYITEKFW 301
DB 227 QAHLLKVDYGR--SHKPLPK-----GTMMETLSRKYFLAFENSLHPDITTEKLM 274
QY 302 R-PLKGVVPVYVGSITDM--LPSNKSALIVSESHPRELASIYRRLDSDRLYEAY 357
```

```

DB      275 RALBEMWAPVVLG-PSSSNYERFLPPD-ARIHVDFFSPDLARVQCELDKHARFLSY 332
CY      358 VEMKXGGEISNORLITLALRER--KMGVDQVNDYDAFECEMCTKWMANIRLQ 409
DB      333 FWR-----ETLRPRFSFWMALD-----FKACMKCQGESRYQ 364

RESULT 8
FUT3, HUMAN
AC P21217; Q99448; Q99449;
DT 01-MAY-1991 (rel. 18, Created)
DT 01-MAY-1991 (rel. 18, Last sequence update)
DT 10-OCT-2003 (rel. 42, Last annotation update)
DE Galactoside 3(4)-L-fucosyltransferase (EC 2.4.1.65) (Blood group Lewis
DE alpha-4-fucosyltransferase) (Lewis Fx) (Fucosyltransferase 3) (FUCT-
DE iiii).
GN FUT3 OR LE OR FT3B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RX SEQUENCE FROM N.A.
RX MEDLINE=91032981; PubMed=1977660;
RA Kukowska-Latallo J.F., Larsen R.D., Nair R.P., Lowe J.B.;
RT "A cloned human cDNA determines expression of a mouse scage-specific
RT embryonic antigen and the Lewis blood group
RT alpha(1,3/1,4) fucosyltransferase."
RL Genes Dev. 4:1288-1303(1990).
RN [12]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95378289; PubMed=7650030;
RA Cameron H.S., Szczepaniak D., Weston W.;
RT "Expression of human chromosome 19p alpha(1,3)-fucosyltransferase
RT genes in normal tissues. Alternative splicing, polyadenylation, and
RT isoforms."
RL J. Biol. Chem. 270:20112-20122(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Squamous cell carcinoma;
RA Rahm I., Schmidt L.R., Wahl D., Drayson E., Maslanik W.,
RA Strahan P.L., Pettiford D.E.;
RT "Isolation and expression of human alpha (1,3/1,4)
RT fucosyltransferase."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP VARIANT LE(-) MET-105.
RX MEDLINE=94059067; PubMed=8240322;
RA Elmgren A., Rydberg L., Larson G.;
RT "Genotypic heterogeneity among Lewis negative individuals."
RL Biochem. Biophys. Res. Commun. 196:515-520(1993).
RN [5]
RP VARIANTS LE(-) ARG-20; SER-170 AND ALA-336.
RX MEDLINE=94059082; PubMed=8240337;
RA Nishihara S., Yazawa S., Iwasaki H., Nakazato M., Kudo T.,
RA Narimatsu H.;
RT "Alpha (1,3/1,4)fucosyltransferase (FUCT-III) gene is inactivated by
RT a single amino acid substitution in Lewis histo-blood type negative
RT individuals."
RL Biochem. Biophys. Res. Commun. 196:624-631(1993).
RN [6]
RP VARIANTS LE(-) ARG-20 AND SER-170.
RX MEDLINE=94033579; PubMed=8219240;
RA Koda Y., Kimura H., Mekada E.;
RT "Analysis of Lewis fucosyltransferase genes from the human gastric
RT mucosa of Lewis-positive and -negative individuals."
RL Blood 82:2915-2919(1993).
RN [7]
RP VARIANTS LE(-) ARG-20 AND LYS-356.
RX MEDLINE=94342559; PubMed=8063716;
RA Mollicone R., Reguigne I., Kelly R.J., Fletcher A., Watt J.,
RA Chaffield S., Aziz A., Cameron H.S., Weston B.W., Lowe J.B., Oriol R.;
RT "Molecular basis for Lewis alpha(1,3/1,4)-fucosyltransferase gene
RT deficiency (FUT3) found in Lewis-negative Indonesian pedigrees."
RL J. Biol. Chem. 269:20987-20994(1994).
RN [8]
RP VARIANT LE(-) LYS-356.
RX MEDLINE=95050753; PubMed=7961897;
RA Nishihara S., Narimatsu H., Iwasaki H., Yazawa S., Akamatsu S.,
RA Ando T., Sato T., Narimatsu H.;
RT "Molecular genetic analysis of the human Lewis histo-blood group
RT system."
RL J. Biol. Chem. 269:29271-29278(1994).
RN [9]
RP VARIANTS LE(-) ARG-20; ARG-68; MET-105 AND LYS-356.
RX MEDLINE=96243526; PubMed=8801770;
RA Elmgren A., Boerjeson C., Svensson L., Rydberg L., Larson G.;
RT "DNA sequencing and screening for point mutations in the human Lewis
RT 'FUT3' gene enables molecular genotyping of the human Lewis blood
RT group system."
RL Vox Sang. 70:97-103(1996).
RN [10]
RP VARIANTS LE(-) ARG-68 AND MET-105.
RX MEDLINE=97413801; PubMed=9268337;
RA Elmgren A., Mollicone R., Costache M., Boerjeson C., Oriol R.,
RA Harrington J., Larson G.;
RT "Significance of individual point mutations, T202C and C314T, in the
RT human Lewis 'FUT3' gene for expression of Lewis antigens by the human
RT alpha(1,3/1,4)-fucosyltransferase, Fuc-TIII."
RL J. Biol. Chem. 272:21994-21998(1997).
RN [11]
RP VARIANTS LE(+) LYS-102 AND ALA-124, AND VARIANTS LE(-) ASN-162;
RP ARG-223 AND MET-270.
RX MEDLINE=98366989; PubMed=9703429;
RA Pang H., Liu Y., Koda Y., Soejima M., Jia J., Schlaphoff T.,
RA du Toit E.D., Kimura H.;
RT "Five novel missense mutations of the Lewis gene 'FUT3' in African
RT 'Xhosa' and Caucasian populations in South Africa."
RL Hum. Genet. 102:675-680(1998).
RN [12]
RP FUNCTION: MAY CATALYZE ALPHA-1,3 AND ALPHA-1,4 GLYCOSIDIC LINKAGES
INVOLVED IN THE EXPRESSION OF VIM-2, LEWIS A, LEWIS B, SIALYL
LEWIS X AND LEWIS X/SSER-1 ANTIGENS. MAY BE INVOLVED IN BLOOD
GROUP LEWIS DETERMINATION; LEWIS-POSITIVE (LE(+)) INDIVIDUALS
HAVE AN ACTIVE ENZYME WHILE LEWIS-NEGATIVE (LE(-)) INDIVIDUALS
HAVE AN INACTIVE ENZYME.
- CATALYTIC ACTIVITY: GDP-beta-L-fucose + beta-D-galactosyl-(1->3)-
- N-acetyl-D-glucosaminyl-R = GDP + beta-D-galactosyl-(1->3)-[alpha-
- L-fucosyl-(1->4)]-N-acetyl-beta-D-glucosaminyl-R.
- PATHWAY: Glycosylation.
- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
form in trans cisternae of Golgi.
- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN STOMACH, COLON, SMALL
INTESTINE, LUNG AND KIDNEY AND TO A LESSER EXTENT IN SALIVARY
GLAND, BLADDER, UTERUS AND LIVER.
- MISCELLANEOUS: Also acts on the corresponding 1,4-galactosyl
derivative, forming 1,3-L-fucosyl links.
- SIMILARITY: Belongs to the fucosyltransferase family 10.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
or send an email to license@ebi-sib.ch).
```

```
DR PIR: A36669; A36669.  
DR Genew: HGNC:4014; FUT3.  
DR MIM: 111100; .  
DR GO: GO:0005624; C:membrane fraction; TAS.  
DR GO: GO:0008417; F:fucosyltransferase activity; TAS.  
DR GO: GO:0005975; P:carbohydrate metabolism; TAS.  
DR InterPro: IPR001503; Glyco trans 10.  
DR Pfam: PF00853; Glyco_transf_10; I.  
DR Transf: Glycosyltransferase; Glycoprotein; Transmembrane;  
KW Signal-anchor; Golgi stack; Polymorphism; Blood group antigen.  
FT DOMAIN 1 15  
FT TRANSMEM 16 34  
FT (POTENTIAL).  
FT LUMENAL, CATALYTIC (POTENTIAL).  
FT N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 154 154  
FT CARBOHYD 185 185  
FT L-N-LINKED (GLCNAC. . .) (PROBABLE).  
FT L-> R (IN LE(-)).  
FT /Ftid-var 003426.  
FT W-> R (IN LE(-)).  
FT /Ftid-var 007959.  
FT /Ftid-var 007960.  
FT Q-> K (IN LE(+)).  
FT /Ftid-var 007960.  
FT T-> M (IN LE(-)).  
FT /Ftid-var 003427.  
FT S-> A (IN LE(+)).  
FT /Ftid-var 007961.  
FT D-> N (IN LE(-)).  
FT /Ftid-var 007962.  
FT G-> S (IN LE(-)); COMPLETELY INACTIVE).  
FT /Ftid-var 003428.  
FT G-> R (IN LE(-)).  
FT /Ftid-var 007963.  
FT V-> M (IN LE(-)).  
FT /Ftid-var 007964.  
FT D-> A (IN LE(-)).  
FT /Ftid-var 003429.  
FT I-> K (IN LE(-)); LESS THAN 10% REDUCTION  
FT IN ACTIVITY).  
FT VARIANT 356 356  
FT VARIANT 356 356  
SQ SEQUENCE 361 AA; 42117 MW; BF4398044F19C284 CRC64;  
Query Match 9.1%; Score 239.5; DB 1; Length 361;  
Best Local Similarity 26.5%; Pred. No. 1,4e-11;  
Matches 94; Conservative 43; Mismatches 115; Indels 103; Gaps 18;  
QY 90 IMLWSPITGFTGLGQC-----GADACFTTNR-----TYLHHMTKAFLYGTDEN 137  
Db 65 ILMWTFPHIPIVA-LSRCSEWVPGADCHITADRKVYPOADIVVH-----WD 112  
QY 138 IDS-----LPLRKAHDMVAVFHEESPKNKYLFHKPVITLFNYTATFSRSHSLPTTQ 191  
Db 113 IMNSPKSRLLPPSPRPGQGRWIMFNLEPPNCCHL--EALDRYFN-----LTMS 158  
QY 192 YLESIEVLSKLRIVP-----LQSKNKLRLRLAPLVYVQSDCPSPDSYVRE 240  
Db 159 YRSDSDIFTPYGLWPSGQPAHPPLNLSAKTEL-----VAMAVSNWKLDSARVRYQ 212  
QY 241 LMTYEVDSYGECLRNKDLPOOLKNPASMDAGFYRIIAQYKFLIAPENACDYTEKEF 300  
Db 213 LQHLAKVDYGR--SHKPLPK-----GIMWELTSRKFLYAFENSLHPYITEKL 260  
QY 301 WR-PLKLGVPYVYGSPTTDW---LPSNKAIIIVSESHRELASIRRLDSDRILYA 356  
Db 261 WRALAEWAVPVVLG--PSRSNRYERPLPD-AFIHVDFOQSPKDLARLYQETDKDHARYLS 318  
QY 357 YVEMKLKGEISNQLLTALRER--KMGVDVQODVYIDAFECWCTKVMANIRLQ 409  
Db 319 YPFMR-----ETLRPRSPFWALD-----FCKACWKLQDSRQ 351
```

```
AC 019058;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Galactoside 3(4)-L-fucosyltransferase (EC 2.4.1.65) (Blood group Lewis  
DE alpha-4-fucosyltransferase) (Lewis Fx) (Fucosyltransferase 3) (FUT3-  
DE IIT) (Alpha-3/4-fucosyltransferase).  
GN FUT3.  
OS Pan troglodytes (Chimpanzee).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
OX NCBI_Taxid:9598;  
RN [1]  
RP SEQUENCE FROM N.A., AND VARIANTS GLY-162 AND MET-304.  
RX MEDLINE:98037800; PubMed:9368041;  
RA Costache M., Apoil P.-A., Calileau A., Elmgren A., Larson G.,  
RA Henry S., Blancher A., Iordachescu D., Oriol R., Mollifone R.;  
RT "Evolution of fucosyltransferase genes in vertebrates";  
RL J. Biol. Chem. 272:29721-29728(1997).  
CC -FUNCTION: May catalyze alpha-1,3 and alpha-1,4 glycosidic linkages  
CC involved in the expression of sialyl Lewis X and Lewis X/SEA-1  
CC antigens. It may be involved in blood group Lewis determination  
CC (By similarity).  
CC -CATALYTIC ACTIVITY: GDP-beta-L-fucose + beta-D-galactosyl-(1->3)-  
CC N-acetyl-D-glucosaminyl-R = GDP + beta-D-galactosyl-(1->3)-[alpha-  
CC L-fucosyl-(1->4)]-N-acetyl-beta-D-glucosaminyl-R.  
CC -PATHWAY: Glycosylation.  
CC -SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
CC form in trans cisternae of Golgi (By similarity).  
CC -POLYMORPHISM: There are two alleles, A and B. Allele A has Arg-  
CC 162 and Val-304. Allele B has Gly-162 and Met-304.  
CC -MISCELLANEOUS: Also acts on the corresponding 1,4-galactosyl  
CC derivative, forming 1,3-L-fucosyl links.  
CC -SIMILARITY: Belongs to the glycosyltransferase family 10.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL: Y14033; CAA74360.1; .  
DR InterPro: IPR001503; Glyco trans 10.  
DR Pfam: PF00852; Glyco_transf_10; I.  
KW Transf: Glycosyltransferase; Glycoprotein; Transmembrane;  
KW Signal-anchor; Golgi stack; Polymorphism.  
FT DOMAIN 1 14  
FT TRANSMEM 15 34  
FT (POTENTIAL).  
FT LUMENAL, CATALYTIC (POTENTIAL).  
FT CARBOHYD 165 165  
FT CARBOHYD 196 196  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 162 162  
FT VARIANT 304 304  
FT R-> G (in allele B).  
FT V-> M (in allele A).  
SQ SEQUENCE 372 AA; 43233 MW; 649CBF8B0A7BD74C CRC64;  
Query Match 8.9%; Score 234.5; DB 1; Length 372;  
Best Local Similarity 26.4%; Pred. No. 3,6e-11;  
Matches 92; Conservative 43; Mismatches 123; Indels 91; Gaps 18;  
QY 90 IMLWSPITGFTGLGQC-----GADACFTTNR-----TYLHHMTKAFLYGTDEN 137  
Db 76 ILMWTFPHIPIVA-LSRCSEWVPGADCHITADRKVYPOADIVVHWD---IMVNSPKSR 131  
QY 138 IDSPLRKAHDMVAVFHEESPKNKYLFHKPVITLFNYTATFSRSHSLPTTQYLESIE 197  
Db 132 LP--PSRPGQGRWIMFNLEPPNCCHL--EALDRYFN-----LTMSYSDSD 175  
QY 198 VLKSLRIYVP-----LQSKNKLRLRLAPLVYVQSDCPSPDSYVRELMTYIE 246  
Db 176 IFTPYGLWPSGQPAHPPLNLSAKTEL-----VAMAVSNWKLDSARVRYQSLQHLK 229
```

| FT         | DOMAIN   | 1   | 55        | CYTOPLASMIC (POTENTIAL).  |
|------------|--|---|-----------|---|
| FT         | DOMAIN   | 56  | 78        | SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)                        |
| FT         | DOMAIN   | 79  | 389       | (POTENTIAL).  |
| FT         | DOMAIN   | 115   | 123       | LUMENAL, CATALYTIC (POTENTIAL).                                 |
| FT         | DISULFID   | 258   | 261       | BY SIMILARITY.  |
| FT         | DISULFID   | 365   | 368       | BY SIMILARITY.  |
| FT         | CARBOHYD   | 128   | 128       | BY SIMILARITY.  |
| FT         | CARBOHYD   | 338   | 338       | N-LINKED (GLCNAC. . .) (POTENTIAL).                             |
| FT         | VARSLLIC   | 1   | 51        | N-LINKED (GLCNAC. . .) (POTENTIAL).                             |
| FT         | SEQUENCE   | 389 AA;   | 44494 MW; | MPKPCPACISLSTGTHRLILPFPDMKAPSMESKEATCNSSS                       |
| FT         | SEQUENCE   | 389 AA;   | 44494 MW; | PGFAPEPTG -> MNCI (in isoform 2).                               |
| FT         | SEQUENCE   | 389 AA;   | 44494 MW; | /PFIID=VSP_001781.  |
| FT         | SEQUENCE   | 389 AA;   | 44494 MW; | 118FC6B2378B99C6 CRC64;   |
| FT         | SEQUENCE   | 389 AA;   | 44494 MW; | 8.9%; Score 233.5; DB 1; Length 389;                            |
| FT         | SEQUENCE   | 389 AA;   | 44494 MW; | Best Local Similarity 25.3%; Pred. No. 4,6e-11;                 |
| FT         | SEQUENCE   | 389 AA;   | 44494 MW; | Matches 87; Conservative 53; Mismatches 116; Indels 75; Gaps 16 |
| QY         | 90   | IMLWSPSLTGTGRLL-----GCCGACACFTTIRTYL-----HHH--MKAPLFTGT     | 134       |   |
| QY         | 96   | ILIMWPFPTNPPPELPGDTCTRYMACSRLSARSLASADAVFTHRELQROSL----     | 151       |   |
| Db         | 135  | DNINDSLPLPKAH-HDMAYVHESSPKNKKLFPHKPYITLFNTATFSESHLPLTYQL    | 193       |   |
| QY         | 152  | -----LELDQRPHQCPVWMASSMSPENTHGL-HR-FRGLFMWVLSYRRDSOILFVPGRL | 203       |   |
| Db         | 194  | ESIEVYLSIRLYVPLQSKNKLRLKELAVLVYVQSDCCPSPDRDSVYELMTYLEVDSYGC | 253       |   |
| QY         | 204  | ELTSGPTS-----PLPAKSPV-----AAWVINSFPOEQRAKLYQLAPHQVDFGSA     | 252       |   |
| Db         | 254  | LNNKULPOLKNKPSMDADGTYRIIAQYKFLIFENNAVCODDYTERKFW-PLKLGVPVY  | 312       |   |
| QY         | 253  | -----SGRPLCNKCLLPLTARIRYLLAFENSGHRDYTERKFMNALLAGAVPA        | 301       |   |
| QY         | 313  | YGSPSIT-DWLPSNKAIIIVSEFSHRELASIYRRLSDDRLYEAVYEWKLKGELSNOR   | 370       |   |
| Db         | 302  | LGPRAITYEAFVPPD-AFVNHDDPSARLAVFL--VSMWESSYRGFPAMRDLRY--R    | 355       |   |
| QY         | 371  | LITALPERKGVQDVQNDYIDAEFCWCTK                                | 401       |   |
| Db         | 356  | LLGDWREK-----FCTICAR  | 370       |   |
| RESULT 11  |  |   |           |   |
| FUI3_ARATH | STANDARD;  | PRT;  | 401 AA.   |   |
| AC         | Q9C8W3; Q8RYC1;  |   |           |   |
| DT         | 16-OCT-2001 (Rel. 40; Created)   |   |           |   |
| DT         | 16-OCT-2001 (Rel. 40; Last sequence update)                            |   |           |   |
| DT         | 28-FEB-2003 (Rel. 41; Last annotation update)                          |   |           |   |
| DE         | Alpha-(1,4)-fucosyltransferase (EC 2.4.1.-) (FT4-M) (Galactoside 3(4)- |   |           |   |
| DE         | L-fucosyltransferase) (PUTC) (ACFUT13).                                |   |           |   |
| GN         | FUT13 OR AT7G71990 OR F17M19.14.                                       |   |           |   |
| OC         | Arabidopsis thaliana (Mouse-ear cress).                                |   |           |   |
| OC         | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;     |   |           |   |
| OC         | Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;   |   |           |   |
| OC         | eumecids II; Brassicales; Brassicaceae; Arabidopsis.                   |   |           |   |
| CX         | NCBI_TaxID=3702;   |   |           |   |
| RN         | [1]  |   |           |   |
| RP         | SEQUENCE FROM N.A.   |   |           |   |
| RP         | STRAIN=cv. Columbia; TISSUE=Root;                                      |   |           |   |
| RC         | MEDLINE=21133304; PubMed=11420147;                                     |   |           |   |
| RA         | Wilson I.B., Rendic D., Freilinger A., Dunic J., Altman F., Mucha J.,  |   |           |   |
| RA         | Muller S., Hauser M.T.;  |   |           |   |
| RT         | "Cloning and expression of cDNAs encoding alpha1,3-fucosyltransferase  |   |           |   |
| RT         | homologues from Arabidopsis thaliana";                                 |   |           |   |
| RL         | Biochim. Biophys. Acta 1527:88-96(2001).                               |   |           |   |
| RN         | [2]  |   |           |   |
| RP         | SEQUENCE FROM N.A.   |   |           |   |
| RP         | STRAIN=cv. Columbia;   |   |           |   |
| RC         | MEDLINE=21016719; PubMed=1130712;                                      |   |           |   |
| RA         | Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,         |   |           |   |
| RA         | White O., Alonso J., Alcafi H., Araujo R., Bowman C.L., Brooks S.Y.,   |   |           |   |

|  |  |   |
|--|--|---|
| Cy   | 170  | VITFENYATFSRSHSLPTLTQYLSEIYLSKLRYLVLFOSKKRLRKLAFLVYQOSI   |
| Dd   | 182  | -----ISHAADVDQTIIYAASGL-FHNNRNHYHSIPKHNDV-----LVYMSSS     |
| Cy   | 226  | DPSPDSDSYRELMNTTIEVDSTGYECCLRN-KDLPOOLK-----NPASMADGFFY   |
| Dd   | 225  | LP--HRDRLAKSLLDLIIPHISFGCKLNNGVGDLSALMYDECVAAEHNAKKVYDIJH |
| Cy   | 279  | AQCYFIAPFNACDDYITEKFMPRLKGVPPVYGSPFITDWLPENKSAIIVSEFSI    |
| Dd   | 283  | SHYKFVALIENTAVESVIYEKLFYALDSGSPIPIFGASNQDVPPH-SVIDSGSRFG  |
| Cy   | 339  | EIASYIRLLSDRLYEAYVEWKLXGEISNQRLTLALRRKMGVQVONDNYIDAFBS    |
| Dd   | 342  | ELAAVYVRLLDDPVAASEYHAMRRCGLMGNVGTKRAVS-----LDTLPD         |
| Cy   | 399  | CTKV 402  |
| Dd   | 388  | CEBI 391  |
| <br>RESULT 12<br>FUT7 HUMAN<br>ID _FUT7_HUMAN STANDARD; PRT; 342 AA. |  |   |
| AC   | 011130:  |   |
| DT   | 01-OCT-1996 (Rel. 34, Last sequence update)  |   |
| DT   | 01-OCT-1996 (Rel. 34, Last sequence update)  |   |
| DT   | 10-OCT-2003 (Rel. 42, Last annotation update)  |   |
| DE   | Alpha-(1,3)-fucosyltransferase (EC 2.4.1.-) (Galactoside 3-L-fucosyltransferase) (Fucosyltransferase 7) (FUCT-VII) (Selectin-ligand synthase). |   |
| DE   | ligand synthase).  |   |
| GN   | FUT7.  |   |
| OC   | Homo sapiens (Human).  |   |
| OC   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  |   |
| OC   | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  |   |
| OX   | NCB I Taxid=9606;  |   |
| OX   | (1)  |   |
| RX   | SEQUENCE FROM N.A.   |   |
| RX   | MEDLINE=94266898; PubMed=8207002;  |   |
| RA   | Natsuka S., Gersten K.M., Zenita K., Kannagi R., Lowe J.B.;  |   |
| RT   | "Molecular cloning of a cDNA encoding a novel human leukocyte  |   |
| RT   | alpha-1,3-fucosyltransferase capable of synthesizing the sialyl  |   |
| RT   | Lewis x determinant".  |   |
| RT   | J. Biol. Chem. 269:16789-16794(1994).  |   |
| RN   | [2]  |   |
| RN   | REVSIONS.  |   |
| RX   | MEDLINE=94327669; PubMed=8051184;  |   |
| RA   | Natsuka S., Gersten K.M., Zenita K., Kannagi R., Lowe J.B.;  |   |
| RA   | "Molecular cloning of a cDNA encoding a novel human leukocyte  |   |
| RA   | alpha-1,3-fucosyltransferase capable of synthesizing the sialyl  |   |
| RA   | Lewis x determinant".  |   |
| RA   | J. Biol. Chem. 269:14730-14737(1994).  |   |
| RN   | [3]  |   |
| RN   | SEQUENCE FROM N.A.   |   |
| RX   | MEDLINE=94237894; PubMed=8182079;  |   |
| RA   | Sasaki K., Kurada K., Funayama K., Nagata M., Watanabe E., Ohta S.,  |   |
| RA   | Hara N., Nishi T.;   |   |
| RA   | "Expression cloning of a novel alpha 1,3-fucosyltransferase that is  |   |
| RA   | involved in biosynthesis of the sialyl Lewis x carbohydrate  |   |
| RA   | determinants in leukocytes."   |   |
| RA   | J. Biol. Chem. 269:14730-14737(1994).  |   |
| RN   | [4]  |   |
| RN   | SEQUENCE FROM N.A.   |   |
| RA   | Hiraiwa N., Hiraiwa M., Kannagi R.;  |   |
| RA   | "The human selectin-I-gand synthase (hFUC-T VII) gene structure and  |   |
| RA   | characterization of the promoter."   |   |
| RA   | Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  |   |
| RP   | [5]  |   |
| RP   | DISULFIDE BONDS.   |   |
| RP   | MEDLINE=21319013; PubMed=11425803;   |   |
| RP   | de Vries T., Yen T.Y., Joshn R.K., Storm J., van Den Bijnden D.H.,   |   |
| RP   | Knegtel R.M.A., Bunschoten H., Joziasse D.H., Macher B.A.;   |   |



"Neighboring cysteine residues in human fucosyltransferase VII are engaged in disulfide bridges, forming small loop structures.";  
 RT Glycobiology 11:423-433 (2001).  
 CC -1- FUNCTION: May catalyze alpha-1,3 glycosidic linkages involved in the expression of sialyl Lewis x antigens.  
 CC -1- CATALYTIC ACTIVITY: GDP-L-fucose + alpha-2,3-Neu-N-acetyl-1,4-beta-D-galactosyl-N-acetyl-1-D-glucosaminyl-R = GDP + alpha-2,3-Neu-N-acetyl-1,4-beta-D-galactosyl-(alpha-1,3-L-fucosyl)-N-acetyl-D-glucosaminyl-R.  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound form in trans cisternae of Golgi.  
 CC -1- TISSUE SPECIFICITY: LEUKOCYTIC/MELOID LINEAGE CELLS.  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 10.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X78031; CAA54962.1; -  
 DR EMBL: U11282; AAA20468.1; -  
 DR EMBL: U08112; AAA56869.1; -  
 DR EMBL: AB012668; AAA32819.1; -  
 DR PIR: A54057; A54057.  
 DR Genew: HGNC:4018; FUT7.  
 DR MIM: 602030; -  
 DR GO: GO:0016021; C:integral to membrane; TAS.  
 DR GO: GO:0008417; F:fucosyltransferase activity; TAS.  
 DR InterPro: IPR001503; Glyco\_transf\_10; 1.  
 DR Pfam: PF00852; Glyco\_transf\_10; 1.  
 DR Transfaser: Glycosyltransferase; Transmembrane; Glycoprotein; KW Signal-anchor; Golgi stack.  
 KM Signal-anchor; Golgi stack.  
 FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 15 36 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).  
 FT DOMAIN 37 342 LUMENAL, CATALYTIC (POTENTIAL).  
 FT DISULFID 68 76  
 FT DISULFID 211 214  
 FT DISULFID 318 321  
 FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 161 162 GP -> A (IN REF. 1; AAA56869).  
 FT CONFLICT 304 305 RL -> SV (IN REF. 1; AAA56869).  
 SQ SEQUENCE 342 AA; 39238 MW; D31BFP9DD64DFAB CRC64;  
 Query Match 8.5%; Score 224; DB 1; Length 342;  
 Best Local Similarity 26.2%; Pred. No. 2.2e-10;  
 Matches 85; Conservative 52; Mismatches 133; Indels 54; Gaps 13;  
 QY 83 WEIDSYF-IMLWSPLTGEGTSL- - - - -GQCADACFFITNRYLHHMT 126  
 DB 31 WLGSAPRTPAPQPTITLIVHWFPTQPELPSDCTGIGARCLANRSL-ASA 88  
 QY 127 KAFLYGIDFNIDSLPLP- - - - -RKAHDAVHESPPKNNYLFKPYITLFNTATSRH 183  
 DB 89 DAVVHHRRLQTRSHLPLAQRPRGQPVWMASSPSHTLSH- - - - -LRGIFMWLSTRRD 146  
 QY 184 SHLPITLYLSIEVLKSLRYLVPLQSKNKLKRLAPLVVQSDCDPPSRDYSVRELMT 243  
 DB 147 SDIFVYVYGLF- - - - -PHMGSPPLPAPSRV- - - - -AAWVVSNGFORQARLYEQ LAP 195  
 QY 244 YIEVDYSGCLNKKLPQOLKPKASMDAGYRIIAQYKFLAEMNAVCDYITEKWR- 302  
 DB 196 HIRVDVFGRA- - - - -NGRPLCASCLVFTVAQYRFLENSQHRDYITEKWRN 244  
 QY 303 PKLGVVPPYVYSPSIT- - - - -DLPSKSAIIVSEFSPRELASYRLSDSDRLYEAVEM 360  
 DB 245 ALVAGIVPVVLPAPPATYAFVAPAD-ATVHVDDPSASALAPLITGM-NSRRQRRFFAW 301

QY 361 KXGEISNORLLTALRRKXGVOD 384  
 DB 302 RDRURV- - - - -RLFTDWRERFCAICD 322  
 RESULT 13  
 ID FUT4 HUMAN STANDARD; PRT; 405 AA.  
 AC P22083;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.1.) (Galactoside 3-L-fucosyltransferase) (Fucosyltransferase 4) (FUT4-IV) (ELAM-1 ligand fucosyltransferase).  
 DE FUT4 OR ELFT.  
 GN Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RX MEDLINE=91373370; PubMed=171630;  
 RA Lowe J.B., Kukowska-Latallo J.F., Nair R.P., Larsen R.D., Marks R.M., Macher B.A., Kelly R.V., Ernst L.K.;  
 RT "Molecular cloning of a human fucosyltransferase gene that determines expression of the Lewis x and VIM-2 epitopes but not ELAM-1-dependent cell adhesion.";  
 RT J. Biol. Chem. 266:17467-17477 (1991).  
 RN [2]  
 RP MEDLINE=91084863; PubMed=1702034;  
 RX Goetz S.E., Hession C., Goff D., Griffiths B., Tizard R., Newman B., Chi-Rosso G., Lobb R.;  
 RA "ELFT: a gene that directs the expression of an ELAM-1 ligand.";  
 RT Cell 63:1349-1356 (1990).  
 RN [3]  
 RP SEQUENCE OF 1-400 FROM N.A.  
 RX MEDLINE=92042084; PubMed=1718983;  
 RA Kumar R., Potvin B., Muller W.A., Stanley P.;  
 RT "Cloning of a human alpha(1,3)-fucosyltransferase gene that encodes ELFT but does not confer ELAM-1 recognition on Chinese hamster ovary cell transfectants.";  
 RT J. Biol. Chem. 266:21777-21783 (1991).  
 CC -1- FUNCTION: May catalyze alpha-1,3 glycosidic linkages involved in the expression of Lewis X/SSeA-1 and VIM-2 antigens.  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound form in trans cisternae of Golgi.  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 10.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M65030; AAA92977.1; -  
 DR EMBL: M58596; AAA63172.1; -  
 DR EMBL: M58597; AAA63173.1; ALT INIT.  
 DR EMBL: S65161; AAB20349.1; -  
 DR PIR: B36340; B36340.  
 DR Genew: HGNC:4015; FUT4.  
 DR MIM: 104230; -  
 DR GO: GO:0005624; C:membrane fraction; TAS.  
 DR GO: GO:0008417; F:fucosyltransferase activity; TAS.  
 DR InterPro: IPR001503; Glyco\_transf\_10.  
 DR Pfam: PF00852; Glyco\_transf\_10; 1.  
 DR Transfaser: Glycosyltransferase; Transmembrane; Glycoprotein; KM

KW Signal-anchor; Golgi stack. CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 1 22 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT TRANSMEM 23 47  
 FT  
 FT DOMAIN 48 405  
 FT CARBOHYD 91 91 LUMENAL, CATALYTIC (POTENTIAL).  
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 241 241 P -> R (IN REF. 2 AND 3).  
 FT SEQUENCE 405 AA, 45569 MW, DE72E1DC390268D CRC64;  
 SC  
 Query Match 8.3%; Score 219; DB 1; Length 405;  
 Best Local Similarity 23.7%; Pred. No. 6.9e-10;  
 Matches 92; Conservative 38; Mismatches 138; Indels 120; Gaps 16;  
 QY 90 IMLWSPITGEGT-----RLGQCGADACFTTINFTYHHNTAFIFYGTFD-- 136  
 DB 66 VLLWMEFPGGRDAPRPPDCCPLRFNISG---CRLLTDRA--SYGEAQAVLFHHRDLVKG 120  
 QY 137 -----NIDSLPL-----PRKAHDMVAFHESPKNNYK 164  
 DB 121 PPDMPPPMGIOANTAEVDIRLIDYEAAAAAALATSSPPRPGQRMVWMMFESS- 176  
 QY 165 LFKHKPVI-----TLFNVTATFSRSHSLPTTQYLSIEVLKSLRYLVPLOSKNKLRRRLA 219  
 DB 177 --HSPGLRSLASLFLFWMTLSYRADG-----DVFPVGYLYRSHRPGDPPGCLA 222  
 QY 220 P-----LVVYQSCDDPPSDRDSVREIMTYIEVDSYSGCLRNKDLPOQLKNPAMDA 271  
 DB 223 PLSRQOGLAVMWVSWDERQARVRYHQLSQHVTVDFGRCGQGPVPE----- 272  
 QY 272 DGEFYIIAQYKFLAFENAVCDYITEKFWR-PLKLGVPVYYG--SPSITDMLPSNKA 328  
 DB 273 IGLHTVARKYFLAFENSGHDYITEKLMRNALLAGAVPVLGDRANRYRFPYPRG-AF 331  
 QY 329 ILVSEFSHRELASIRRLDSDDRLYEAVENKLGSELSNQLTALREKMGVQDVND 388  
 DB 332 IHVDDPPSASLSLSTYLLFDNRNPAYRRYFHRRSYAVH-----ITSFWDPM----- 379  
 QY 389 NYIDAFECMCTKVMANIRLOEKGKLPK 416  
 DB 380 -----CRVCOAV-----QRAQDRPK 394  
 RESULT 14  
 ID PUT4 MOUSE STANDARD; PRT: 433 AA.  
 AC 011127;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Alpha-(1,3)-Fucosyltransferase (EC 2.4.1.-) (Galactoside 3-L-fucosyltransferase) (Fucosyltransferase 4) (FUCT-IV).  
 GN PUT4 OR ELFT.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN  
 RP  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
 RC STRAIN=129/Sv; TISSUE=Liver;  
 RX MEDLINE=97037075; Pubmed=8882722;  
 RA Ozawa M., Muramatsu T.;  
 RT "Molecular cloning and expression of a mouse alpha-1,3

RT fucosyltransferase gene that shows homology with the human alpha-1,3 fucosyltransferase IV gene."  
 RL J. Biochem. 119:302-308 (1996).  
 CC -1- FUNCTION: May catalyze alpha-1,3 glycosidic linkages involved in the expression of Lewis X/SSA-1 and VIM-2 antigens.  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound form in trans cisternae of Golgi.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Name=Long;  
 CC IsoId=Q11127-1; Sequence=displayed;  
 CC Name=Short;  
 CC IsoId=Q11127-2; Sequence=VSP 001778;  
 CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN STOMACH AND COLON. IT ALSO EXPRESSED IN THE LUNG, TESTIS, UTERUS, SMALL INTESTINE AND TO A LESSER EXTENT IN SPLEEN, AND OVARY. PRESENT IN TRACE AMOUNTS IN BRAIN, THYMUS, HEART, SMOOTH MUSCLE, KIDNEY AND BONE MARROW. NOT FOUND IN LIVER, SALIVARY GLAND AND PANCREAS.  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 10.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U33457; AAC52269.1; -;  
 DR EMBL; D63380; BAA09697.1; -;  
 DR EMBL; D63379; BAA09696.1; -;  
 DR PIR; A57596; A57596.  
 DR MGD; MGI:95594; Put4.  
 DR InterPro; IPR001503; Glyco trans\_10.  
 DR Pfam; PF00852; Glyco transf\_10; 1.  
 KW Transferase; Glycosyltransferase; Transmembrane; Glycoprotein; Signal-anchor; Golgi stack; Alternative splicing.  
 FT DOMAIN 1 52 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 53 74 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).  
 FT DOMAIN 75 433 LUMENAL, CATALYTIC (POTENTIAL).  
 FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 1 33 Missing (in isoform short).  
 FT FTID=VSP 001778.  
 FT CONFLICT 252 252 O -> P (IN REF. 2).  
 FT CONFLICT 257 257 R -> Q (IN REF. 2).  
 FT CONFLICT 260 260 V -> E (IN REF. 2).  
 FT CONFLICT 273 273 R -> Q (IN REF. 2).  
 SQ SEQUENCE 433 AA; 49481 MW; 2401822F02B5D021 CRC64;  
 Query Match 8.1%; Score 213; DB 1; Length 433;  
 Best Local Similarity 23.4%; Pred. No. 2.3e-09;  
 Matches 89; Conservative 46; Mismatches 140; Indels 104; Gaps 15;  
 QY 90 IMLWSPITGEGT-----RLGQCG-ADACFTTINFTYHHN----- 124  
 DB 92 VLLWMEFPGGRGGRPKSPDCCPLRFNISGCRLLTDRAAAGEAQAVLFHHRDLVKEIDWP 151  
 QY 125 -----MKRALFYSTDENIDSLPLPKA-----HHDMVAFHESPKNNYKLFH 167  
 DB 152 PPKGARERTKALVLRVFDQSGAVITLGALETVSRRPGQRMVWMMFESS-----H 205  
 QY 168 KPVI-----TLFNVTATFSRSHSLPTTQYLSIEVLKSLRYLVPLOSKNKLRRRLA- 220  
 DB 206 TPGRLGAKOLFWMTLSYRTSDVFPVGYLRS-----DPTEGSGGLQPLARRK 257  
 QY 221 --LVVYQSCDDPPDRDSYREIMTYIEVDSYSGCLRNKDLPOQLKNPAMDADEFYRI 278  
 DB 258 GLVAMVAVSNMNEHQARVRYHQLSRHVSVDVFERGTPGPVPE-----AIGLTLTV 307  
 QY 279 AQYKFLAFENAVCDYITEKFWRPLKLGVPVYYG--SPSITDMLPSNKAIIVSEFS 335

```

Db      308 AAYKYVLAPEMSRHYDYITTEKLMWNAFLAGVAVLPPDRANERFPRG-AFIHVDDFP 366
Cc      336 HPRLEIASYIRLDSDDRLYEAVYEWKLGKGEISNORLTALRERKVGQVQNYIDAE 395
Cc      367 NAASLAALLLDLNRVAVYRRYFRMRSPFVH-----ITSFWDGOW----- 407
Cc      396 CMVCTKWANIRLOEKGLPPK 416
Cc      408 CRTCAV-----QTSQDPK 422

RESULT 15
ID      FUL1_ARATH STANDARD; PRT; 501 AA.
AC      Q9LJK1;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Glycoprotein 3-alpha-L-fucosyltransferase A (FC 2.4.1.214) (Core
DE      alpha (1,3)-fucosyltransferase) (Fuc-T C3) (FUCTA) (FUCT1) (ACFUT11).
GN      FUT11 OR A13G19280 OR WY11.20.
OC      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX      NCBI_Taxid=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Columbia; TISSUE=Root;
RX      MEDLINE=21313304; Pubmed=11420147;
RA      Wilson I.B., Rendic D., Freilinger A., Dunic J., Altman F., Mucha J.,
RA      Muller S., Hauser M.T.;
RT      "Cloning and expression of cDNAs encoding alpha1,3-fucosyltransferase
RT      homologues from Arabidopsis thaliana."
RL      Biochim. Biophys. Acta 1527:88-96(2001).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Columbia; TISSUE=Stilques;
RA      Baker H.;
RT      "Molecular cloning and characterization of a plant (Beta vulgaris)
RT      alpha1,4-fucosyltransferase with specificity for Lewis and Lewisb
RT      synthests."
RL      Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Columbia;
RX      MEDLINE=20363099; Pubmed=10907853;
RA      Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RT      "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT      Sequence features of the 4,251,695 bp regions covered by 90 pl, TAC
RT      and BAC clones."
RL      DNA Res. 7:217-221(2000).
RN      [4]
RP      SEQUENCE OF 26-287 FROM N.A.
RC      STRAIN=cv. Wassilewskija;
RA      Kiefer-Meyer M.-C., Faye L., Gomord V.;
RL      Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
Cc      -1- FUNCTION: Involved in cell wall synthesis. Catalyzes alpha-1,3
Cc      glycosidic linkages.
Cc      -1- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-[N-acetyl-beta-D-
Cc      glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-
Cc      glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-
Cc      >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-
Cc      glucosaminyl]asparagine = GDP + N(4)-[N-acetyl-beta-D-
Cc      glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-
Cc      glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-
Cc      >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1->3)]-
Cc      N-acetyl-beta-D-glucosaminyl]asparagine.
Cc      -1- COFACTOR: Manganese.
Cc      -1- PATHWAY: Glycosylation.
Cc      -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
Cc      form in trans cisternae of Golgi (By similarity).

```

```

Cc      -1- SIMILARITY: Belongs to the glycosyltransferase family 10.
Cc      -----
Cc      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
Cc      between the Swiss Institute of Bioinformatics and the EMBL outstation -
Cc      the European Bioinformatics Institute. There are no restrictions on its
Cc      use by non-profit institutions as long as its content is in no way
Cc      modified and this statement is not removed. Usage by and for commercial
Cc      entities requires a license agreement (See http://www.isb-sib.ch/announce/
Cc      or send an email to license@isb-sib.ch).
Cc      -----
DR      EMBL; AP000419; BAB02969.1; -
DR      EMBL; AJ345084; CAC78979.1; -
DR      EMBL; AJ404860; CAC38048.1; -
DR      EMBL; AF277228; AAM68912.1; -
DR      EMBL; AF277229; AAM68913.1; -
DR      InterPro; IPR001503; Glyco_transf_10.
DR      Pfam; PF00852; Glyco_transf_10; 1.
KW      Transferase; Glycosyltransferase; Transmembrane; Glycoprotein;
KW      Signal-anchor; Golgi stack; Cell wall.
FT      DOMAIN 1 39
FT      TRANSMM 40 60
FT      TRANSMM 61 501
FT      DOMAIN 26 35
FT      POLY-SER.
FT      CARBOHYD 64 64
FT      CARBOHYD 337 337
FT      CARBOHYD 420 420
FT      CARBOHYD 481 481
FT      CONFLICT 210 210
SQ      SEQUENCE 501 AA; 56205 MW; 2CEAC85CD1A3DASB CRC64;
Query Match 8.0%; Score 209; DB 1; Length 501;
Best Local Similarity 29.4%; Pred. No. 5.8e-09;
Matches 62; Conservative 35; Mismatches 84; Indels 30; Gaps 6;
Cc      205 LVPIQSNKKRKRIAPLYVYQSDCDPPSDSDSYRELM-TYIVDSYSGSLRKXDPOL 263
Cc      201 MAPQPKTE--KALAAAFI--SNCAKNFRLOALELMKTNVIXIDYGGCHRRD----- 251
Cc      264 KNPASMDADGFYRIAQYKFLAPENAVCDYITEKFWREPLKGVVPVYVYSGPSITDMLP 323
Cc      252 ---GSVER--VEALKRYKESLAFENTNEDYITEKFRQSLVAGSPVYVGAFTIEFPAP 305
Cc      324 SNKAIAIVSEFSHRELASYIRLSDDDLRYEAVYEWKLGKGEISNORLTALRERKVGQ 383
Cc      306 SPDSFLHIKQMDVPAKAVAKKKMYLADNPDAYNQTLRWKHEGSPDSFKALI----- 355
Cc      384 DVQNDYIDAEFCVCTKWANIRLOEKGLP 414
Cc      356 ---DMAAVHSSCRCLFVATRIREQDEKSP 382

```

Search completed: September 9, 2004, 21:41:48  
 Job time : 27 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 21:46:17; Search time 131 Seconds

(without alignments)  
1189.734 Million cell updates/sec

Title: US-10-080-960-2

Perfect score: 2625

Sequence: 1 MKVTGPGQVTDMSQCFNDQ.....LVDRNPFSSQFHWGLVFPD 486

Scoring table: BIOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCR\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCRUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | ID                  | Description        |
|------------|--------|-------------|--------|---------------------|--------------------|
| 1          | 2625   | 100.0       | 486    | US-10-080-960-2     | Sequence 2, Appl1  |
| 2          | 1957.5 | 74.6        | 406    | US-09-731-872-284   | Sequence 284, App  |
| 3          | 1957.5 | 74.6        | 406    | US-09-876-997-284   | Sequence 284, App  |
| 4          | 1015   | 38.7        | 197    | US-09-864-408A-7106 | Sequence 7106, Ap  |
| 5          | 872    | 33.2        | 492    | US-09-844-948-2     | Sequence 2, Appl1  |
| 6          | 872    | 33.2        | 492    | US-10-184-648-20    | Sequence 20, Appl1 |
| 7          | 554.5  | 21.1        | 168    | US-10-467-595-7     | Sequence 7, Appl1  |
| 8          | 551    | 19.8        | 107    | US-09-764-877-1210  | Sequence 1210, Ap  |
| 9          | 521    | 19.8        | 107    | US-10-242-515-1210  | Sequence 1210, Ap  |
| 10         | 487.5  | 18.6        | 261    | US-10-072-877-21    | Sequence 21, Appl1 |
| 11         | 369    | 14.1        | 178    | US-10-080-960-17    | Sequence 17, Appl1 |
| 12         | 318.5  | 12.1        | 157    | US-10-108-260A-3173 | Sequence 3173, Ap  |
| 13         | 282.5  | 10.8        | 365    | US-10-120-319-9     | Sequence 9, Appl1  |
| 14         | 282.5  | 10.8        | 365    | US-10-189-977-9     | Sequence 9, Appl1  |
| 15         | 277.5  | 10.6        | 365    | US-10-392-098-9     | Sequence 9, Appl1  |

|    |       |      |     |    |                      |                    |
|----|-------|------|-----|----|----------------------|--------------------|
| 16 | 263   | 10.0 | 359 | 9  | US-09-863-475A-14    | Sequence 14, Appl1 |
| 17 | 263   | 10.0 | 359 | 13 | US-10-120-319-10     | Sequence 10, Appl1 |
| 18 | 263   | 10.0 | 359 | 14 | US-10-189-977-10     | Sequence 10, Appl1 |
| 19 | 263   | 10.0 | 359 | 14 | US-10-392-098-10     | Sequence 6, Appl1  |
| 20 | 262.5 | 10.0 | 358 | 9  | US-09-733-524-6      | Sequence 5, Appl1  |
| 21 | 266.5 | 9.8  | 502 | 10 | US-09-844-948-5      | Sequence 23, Appl1 |
| 22 | 266.5 | 9.8  | 502 | 12 | US-10-184-648-23     | Sequence 16, Appl1 |
| 23 | 266.5 | 9.7  | 502 | 13 | US-10-080-960-16     | Sequence 4748, Ap  |
| 24 | 255.5 | 9.7  | 336 | 15 | US-10-108-260A-7748  | Sequence 4, Appl1  |
| 25 | 252   | 9.6  | 222 | 10 | US-09-844-948-4      | Sequence 22, Appl1 |
| 26 | 252   | 9.6  | 222 | 12 | US-10-184-648-22     | Sequence 12, Appl1 |
| 27 | 251   | 9.6  | 356 | 13 | US-10-120-319-12     | Sequence 12, Appl1 |
| 28 | 251   | 9.6  | 356 | 14 | US-10-189-977-12     | Sequence 18, Appl1 |
| 29 | 248   | 9.4  | 354 | 9  | US-09-733-524-5      | Sequence 11, Appl1 |
| 30 | 245.5 | 9.4  | 374 | 9  | US-09-863-475A-11    | Sequence 108481,   |
| 31 | 245   | 9.3  | 414 | 16 | US-10-437-963-108481 | Sequence 2, Appl1  |
| 32 | 239.5 | 9.1  | 361 | 9  | US-09-863-475A-2     | Sequence 54364, A  |
| 33 | 238   | 9.1  | 457 | 12 | US-10-425-114-54364  | Sequence 43063, A  |
| 34 | 238   | 9.1  | 450 | 12 | US-10-425-114-43063  | Sequence 12, Appl1 |
| 35 | 237.5 | 9.0  | 355 | 14 | US-10-392-098-12     | Sequence 18, Appl1 |
| 36 | 236   | 9.0  | 139 | 13 | US-10-080-960-18     | Sequence 2, Appl1  |
| 37 | 233.5 | 8.9  | 393 | 9  | US-09-784-077-2      | Sequence 8, Appl1  |
| 38 | 233.5 | 8.9  | 359 | 12 | US-10-700-505-2      | Sequence 11, Appl1 |
| 39 | 232.5 | 8.9  | 355 | 9  | US-09-733-524-8      | Sequence 13, Appl1 |
| 40 | 221   | 8.4  | 405 | 14 | US-10-234-041-11     | Sequence 8, Appl1  |
| 41 | 221   | 8.4  | 530 | 14 | US-10-234-041-13     | Sequence 7, Appl1  |
| 42 | 219   | 8.3  | 405 | 9  | US-09-863-475A-8     | Sequence 11, Appl1 |
| 43 | 216.5 | 8.2  | 432 | 9  | US-09-733-524-7      | Sequence 11, Appl1 |
| 44 | 213   | 8.1  | 433 | 13 | US-10-120-319-11     | Sequence 11, Appl1 |
| 45 | 213   | 8.1  | 433 | 14 | US-10-189-977-11     | Sequence 11, Appl1 |

#### ALIGNMENTS

RESULT 1  
US-10-080-960-2  
Sequence 2, Application US/10080960  
Publication No. US20020197695A1  
GENERAL INFORMATION:  
APPLICANT: Millennium Pharmaceuticals, Inc.  
APPLICANT: Glucksmann, Maria  
APPLICANT: Meyers, Rachel  
TITLE OF INVENTION: 80090, 52874, 52880, 63497, AND 33425  
TITLE OF INVENTION: METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES THEREOF  
FILE REFERENCE: 38155-20044.00  
CURRENT APPLICATION NUMBER: US/10/080,960  
CURRENT FILING DATE: 2001-10-19  
PRIOR APPLICATION NUMBER: US 60/242,040  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: US 60/242,038  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: US 60/241,992  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: US 60/242,637  
PRIOR FILING DATE: 2000-10-23  
NUMBER OF SEQ ID NOS: 37  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 486  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-080-960-2  
Query Match 100.0%; Score 2625; DB 13; Length 486;  
Best Local Similarity 100.0%; Pred. No. 1.4e-247;  
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKVTGPGQVTDMSQCFNDQPLNSTRSSHIKRVVWELGKFERKPKSSLDGGTKME 60  
Db 1 MKVTGPGQVTDMSQCFNDQPLNSTRSSHIKRVVWELGKFERKPKSSLDGGTKME 60  
QY 61 BAPTHINSFLKKEGLTFNRKRWELDSYPTLMWSPLTGTGRLGGCAGACFTINRTY 120

Db 61 EAPTHNSFLKKEGLTFNRRKRWELDSYPLMWSPLTGETRLGQCGADACFTINRTY 120  
QY 121 LHHMTKAFLEFYGTDFNIDSLPLPRKAHDMWAFHEESPKNNYKLFHKPVITLFNYTATF 180  
Db 121 LHHMTKAFLEFYGTDFNIDSLPLPRKAHDMWAFHEESPKNNYKLFHKPVITLFNYTATF 180  
QY 181 SRSHSLPLTTOYLESIEVLKSLRYLVLQSKNKLRLKRLAPLVYVQSDCDPSDRDSYRE 240  
Db 181 SRSHSLPLTTOYLESIEVLKSLRYLVLQSKNKLRLKRLAPLVYVQSDCDPSDRDSYRE 240  
QY 241 LMTYIEVDSYGECIRNKDLPQOLKNPASMDADGFYRIIAQYKFIILAFENAVCDYITEKF 300  
Db 241 LMTYIEVDSYGECIRNKDLPQOLKNPASMDADGFYRIIAQYKFIILAFENAVCDYITEKF 300  
QY 301 WRPLKLVVPVYVYGSPTITDMLPSNKSALIVSEFSHPRELASYIRLSDSDRLYEAYVEM 360  
Db 301 WRPLKLVVPVYVYGSPTITDMLPSNKSALIVSEFSHPRELASYIRLSDSDRLYEAYVEM 360  
QY 361 KLKGEISNORLLTALBERKMGVQDVNODNYIDAFECMTCTKVMANIRLOEKGLPRKWEA 420  
Db 361 KLKGEISNORLLTALBERKMGVQDVNODNYIDAFECMTCTKVMANIRLOEKGLPRKWEA 420  
QY 421 EDTHLSCPEPTVAFSPRLTPPLSLREMTISSFEQSKKEAOLRWLVDRNQNPSQEFM 480  
Db 421 EDTHLSCPEPTVAFSPRLTPPLSLREMTISSFEQSKKEAOLRWLVDRNQNPSQEFM 480  
QY 481 GLVFKD 486  
Db 481 GLVFKD 486  
RESULT 2  
US-09-731-872-284  
; Sequence 284, Application US/09731872  
; Patent No. US20020102604A1  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, Jean Baptiste  
; APPLICANT: Bougueleret, Lydie  
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS  
; FILE REFERENCE: 78. US3. REG  
; CURRENT APPLICATION NUMBER: US/09/731.872  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/169,629  
; PRIOR FILING DATE: 1999-12-08  
; PRIOR APPLICATION NUMBER: US 60/187,470  
; PRIOR FILING DATE: 2000-03-06  
; NUMBER OF SEQ ID NOS: 482  
; SOFTWARE: Patent.pm  
; SEQ ID NO 284  
; LENGTH: 406  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -31...-1  
US-09-731-872-284  
Query Match 74.6%; Score 1957.5; DB 9; Length 406;  
Best Local Similarity 97.9%; Pred. No. 2,2e-182;  
Matches 366; Conservative 1; Mismatches 2; Indels 5; Gaps 1;  
QY 1 MKVTGPPQGVTDMSQCFNDQWPLSNTRSSSEHIKEVMEYELGFKERKEFKSSSLQDGHITKXE 60  
Db 36 MKVTGPPQGVTDMSQCFNDQWPLSNTRSSSEHIKEVMEYELGFKERKEFKSSSLQDGHITKXE 60  
QY 61 EAPTHNSFLKKEGLTFNRRKRWELDSYPLMWSPLTGETRLGQCGADACFTINRTY 120  
Db 61 EAPTHNSFLKKEGLTFNRRKRWELDSYPLMWSPLTGETRLGQCGADACFTINRTY 120  
QY 96 EAPTHNSFLKKEGLTFNRRKRWELDSYPLMWSPLTGETRLGQCGADACFTINRTY 155  
Db 96 EAPTHNSFLKKEGLTFNRRKRWELDSYPLMWSPLTGETRLGQCGADACFTINRTY 155  
QY 121 LHHMTKAFLEFYGTDFNIDSLPLPRKAHDMWAFHEESPKNNYKLFHKPVITLFNYTATF 180  
Db 121 LHHMTKAFLEFYGTDFNIDSLPLPRKAHDMWAFHEESPKNNYKLFHKPVITLFNYTATF 180

Db 156 LHHMTKAFLEFYGTDFNIDSLPLPRKAHDMWAFHEESPKNNYKLFHKPVITLFNYTATF 215  
QY 181 SRSHSLPLTTOYLESIEVLKSLRYLVLQSKNKLRLKRLAPLVYVQSDCDPSDRDSYRE 240  
Db 216 SRSHSLPLTTOYLESIEVLKSLRYLVLQSKNKLRLKRLAPLVYVQSDCDPSDRDSYRE 275  
QY 241 LMTYIEVDSYGECIRNKDLPQOLKNPASMDADGFYRIIAQYKFIILAFENAVCDYITEKF 300  
Db 276 LMTYIEVDSYGECIRNKDLPQOLKNPASMDADGFYRIIAQYKFIILAFENAVCDYITEKF 335  
QY 301 WRPLKLVVPVYVYGSPTITDMLPSNKSALIVSEFSHPRELASYIRLSDSDRLYEAYVEM 360  
Db 301 WRPLKLVVPVYVYGSPTITDMLPSNKSALIVSEFSHPRELASYIRLSDSDRLYEAYVEM 360  
QY 361 KLKGEISNORLLTAL 374  
Db 361 KLKGEISNORLLTAL 374  
RESULT 3  
US-09-876-997-284  
; Sequence 284, Application US/09876997  
; Publication No. US20030152921A1  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, Jean Baptiste  
; APPLICANT: Bougueleret, Lydie  
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS  
; FILE REFERENCE: 78. US4. CIP  
; CURRENT APPLICATION NUMBER: US/09/876.997  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 09/731.872  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/187,470  
; PRIOR FILING DATE: 2000-03-06  
; PRIOR APPLICATION NUMBER: US 60/169,629  
; PRIOR FILING DATE: 1999-12-08  
; NUMBER OF SEQ ID NOS: 482  
; SOFTWARE: Patent.pm  
; SEQ ID NO 284  
; LENGTH: 406  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -31...-1  
US-09-876-997-284  
Query Match 74.6%; Score 1957.5; DB 10; Length 406;  
Best Local Similarity 97.9%; Pred. No. 2,2e-182;  
Matches 366; Conservative 1; Mismatches 2; Indels 5; Gaps 1;  
QY 1 MKVTGPPQGVTDMSQCFNDQWPLSNTRSSSEHIKEVMEYELGFKERKEFKSSSLQDGHITKXE 60  
Db 36 MKVTGPPQGVTDMSQCFNDQWPLSNTRSSSEHIKEVMEYELGFKERKEFKSSSLQDGHITKXE 60  
QY 61 EAPTHNSFLKKEGLTFNRRKRWELDSYPLMWSPLTGETRLGQCGADACFTINRTY 120  
Db 61 EAPTHNSFLKKEGLTFNRRKRWELDSYPLMWSPLTGETRLGQCGADACFTINRTY 120  
QY 96 EAPTHNSFLKKEGLTFNRRKRWELDSYPLMWSPLTGETRLGQCGADACFTINRTY 155  
Db 96 EAPTHNSFLKKEGLTFNRRKRWELDSYPLMWSPLTGETRLGQCGADACFTINRTY 155  
QY 121 LHHMTKAFLEFYGTDFNIDSLPLPRKAHDMWAFHEESPKNNYKLFHKPVITLFNYTATF 180  
Db 121 LHHMTKAFLEFYGTDFNIDSLPLPRKAHDMWAFHEESPKNNYKLFHKPVITLFNYTATF 180  
QY 181 SRSHSLPLTTOYLESIEVLKSLRYLVLQSKNKLRLKRLAPLVYVQSDCDPSDRDSYRE 240  
Db 181 SRSHSLPLTTOYLESIEVLKSLRYLVLQSKNKLRLKRLAPLVYVQSDCDPSDRDSYRE 240  
QY 241 LMTYIEVDSYGECIRNKDLPQOLKNPASMDADGFYRIIAQYKFIILAFENAVCDYITEKF 300  
Db 241 LMTYIEVDSYGECIRNKDLPQOLKNPASMDADGFYRIIAQYKFIILAFENAVCDYITEKF 300  
QY 276 LMTYIEVDSYGECIRNKDLPQOLKNPASMDADGFYRIIAQYKFIILAFENAVCDYITEKF 335  
Db 276 LMTYIEVDSYGECIRNKDLPQOLKNPASMDADGFYRIIAQYKFIILAFENAVCDYITEKF 335  
QY 301 WRPLKLVVPVYVYGSPTITDMLPSNKSALIVSEFSHPRELASYIRLSDSDRLYEAYVEM 360  
Db 301 WRPLKLVVPVYVYGSPTITDMLPSNKSALIVSEFSHPRELASYIRLSDSDRLYEAYVEM 360

Db 336 WRPLKGVVYVYSSSITDMLPSNKSALVSEFSPRELASIRRLDSDRLYAVEM 395  
QY 361 KXGEISNORLTA 374  
Db 396 KXKG-----RSLTS 404

## RESULT 4

US-09-864-408A-7106  
Sequence 7106, Application US/09864408A  
Publication No. US20040009474A1

## GENERAL INFORMATION:

APPLICANT: Leach, Martin D.  
APPLICANT: Shimkets, Richard A.  
TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encd  
FILE REFERENCE: 21402-012  
CURRENT APPLICATION NUMBER: US/09/864,408A  
CURRENT FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: 60/206,690  
PRIOR FILING DATE: 2000-05-24  
NUMBER OF SEQ ID NOS: 9068  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7106  
LENGTH: 197  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-864-408A-7106

Query Match 38.7%; Score 1015; DB 11; Length 197;

Best Local Similarity 99.0%; Pred. No. 1,2e-90;  
Matches 193; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 159 PKNYKLFHKVITLFTNTATPSRSHSLPLTTQYLSIEVLKSLRYLPLQSNKXKRL 218  
Db 3 PENNYKLFHKVITLFTNTATPSRSHSLPLTTQYLSIEVLKSLRYLPLQSNKXKRL 62  
QY 219 APLVYQSDCDPSPDRSDYVRELMTYIEVDSYGECLRNKDLPOOLKNPASMADGFRIT 278  
Db 63 APLVYQSDCDPSPDRSDYVRELMTYIEVDSYGECLRNKDLPOOLKNPASMADGFRIT 122  
QY 279 AQQKFLAFENAVCDYITEKFRPLKLGVPYVYSGPSTIDMLPSNKSALVSEFSHR 338  
Db 123 AQQKFLAFENAVCDYITEKFRPLKLGVPYVYSGPSTIDMLPSNKSALVSEFSHR 182  
QY 339 ELASYIRRLDSDRL 353  
Db 183 ELASYIRRLDSDRL 197

## RESULT 5

US-09-844-948-2  
Sequence 2, Application US/09844948  
Publication No. US20030119161A1

## GENERAL INFORMATION:

APPLICANT: Meyers, Rachel A.  
APPLICANT: Williamson, Mark  
TITLE OF INVENTION: 32132, A NOVEL FUCOSYLTRANSFERASE FAMILY  
FILE REFERENCE: 10448-048001  
CURRENT APPLICATION NUMBER: US/09/844,948  
CURRENT FILING DATE: 2001-04-24  
PRIOR APPLICATION NUMBER: US 60/200,604  
PRIOR FILING DATE: 2000-04-28  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 492  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-844-948-2

Query Match 33.2%; Score 872; DB 10; Length 492;

Best Local Similarity 42.6%; Pred. No. 4.8e-76;  
Matches 118; Conservative 74; Mismatches 146; Indels 20; Gaps 9;

QY 79 EKKWELSDYPIMLWSP-----LTGSTRLOCCGADACFTINTYLAHMTATFLYX 133  
Db 66 RRGREACDLPVILWVSPGLFPHFPDSERI-ECARGCVASRRRALRDSRTRLLYG 124  
QY 134 TDFNIDSLP-PRKXHDWAVFEESPKNNYKLFHKVITLFTNTATPSRSHSLPLTQYL 193  
Db 125 TDFRASAALPLRLAHSMWLLHEESPLNNFLSHGPGIRLFTLSTFSPRSDYPLSLQWL 184  
QY 194 ESTEIVKSLRYVLPLOSNNKLRK-LAPLVYQSDCDPSPDRSDYVRELMTYIEVDSYGE 252  
Db 185 PGYAYLR--RPVPPMERAEWRRRGYAPLLYQSHCDVADRDYRSLMRIIPDSYK 242  
QY 253 CLANKLCP-QOLKNPASMADG--FYRTAQYKFLAFENAVCDYITEKFRPLKGV 309  
Db 243 CLONREPLTRLDYTTATTEDEBLAFSRKFLALEMNCNMYEKLMPYHLGAV 302  
QY 310 PYYGSPSITDMLPSNKSALVSEFSHPRELASIRRLDSDRLYAVEMKXGEISNQ 369  
Db 303 PYYGSPSVADMVNNHSHYLLIDFESPOKLAEIFDLKNDSEYMKYLAHQPGITNQ 362  
QY 370 RLTLALBERKGVQDVNQNYIDAPECWCCTVMANIRLQE-----KGLPP--KSMWED 422  
Db 363 FLIDSLKHRMGVNDLLPVLNYLNGFCVCDYELARLDAEKHAASPGDSPVFEHIAQ 422  
QY 423 TSLSCPEPTVFAFSPTRTPPLSLREMTISFEQSKKEAQLRWIVDRNQNSQEFW 480  
Db 423 SHDCPYPTGFGVNEIPENDSEKEMWLDQYWGQGEALTYAMINNET-EQTKFW 479

## RESULT 6

US-10-184-648-20  
Sequence 20, Application US/10184648  
Publication No. US20030224376A1

## GENERAL INFORMATION:

APPLICANT: Meyers, Rachel E.  
APPLICANT: Williamson, Mark  
APPLICANT: Tsai, Fong-Ying  
APPLICANT: Hunter, John J.  
APPLICANT: MacBeth, Kyle J.  
APPLICANT: Rudolph-Owen, Laura A.  
APPLICANT: Leiby, Kevin R.  
APPLICANT: Kapeller-Libermann, Rosana  
APPLICANT: Olad, Peter U.  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE FAMILY MEMBERS AND USES THEREOF  
FILE REFERENCE: 10448-192001  
CURRENT APPLICATION NUMBER: US/10/184,648  
CURRENT FILING DATE: 2002-06-27  
PRIOR APPLICATION NUMBER: US 09/815,028  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: PCT/US01/09358  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: US 60/191,964  
PRIOR FILING DATE: 2000-03-24  
PRIOR APPLICATION NUMBER: US 09/801,220  
PRIOR FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: PCT/US01/07269  
PRIOR FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: US 60/187,456  
PRIOR FILING DATE: 2000-03-07  
PRIOR APPLICATION NUMBER: US 09/816,714  
PRIOR FILING DATE: 2001-03-23  
PRIOR APPLICATION NUMBER: PCT/US01/09468  
PRIOR FILING DATE: 2001-03-23  
PRIOR APPLICATION NUMBER: US 60/191,865  
PRIOR FILING DATE: 2000-03-24  
PRIOR APPLICATION NUMBER: US 09/844,948  
PRIOR FILING DATE: 2001-04-27  
PRIOR APPLICATION NUMBER: PCT/US01/13805  
PRIOR FILING DATE: 2001-04-27  
PRIOR APPLICATION NUMBER: US 60/200,604

```

; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 09/861,164
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: PCT/US01/16292
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,408
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/863,060
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19138
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/212,079
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/962,678
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: PCT/US01/29963
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/235,044
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 09/973,457
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US 60/238,849
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 10/072,285
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: PCT/US02/03736
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/267,494
; PRIOR FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: US 09/817,910
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: PCT/US01/09633
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/192,092
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 09/842,528
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: PCT/US01/40607
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/199,500
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/882,836
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19543
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/211,730
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/882,872
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19153
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/212,077
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-184-648-20
```

```

Query Match      33.2%; Score 872; DB 12; Length 492;
Best Local Similarity 42.6%; Pred. No. 4,8e-76;
Matches 178; Conservative 74; Mismatches 146; Indels 20; Gaps 9;
```

```

QY 79 RKRKELDSYPIIMWSP-----LTGRTGLGCGGADACFFITNRYTLHHMKATLFYG 133
DB 66 RPRREHAGDLPVLLWMSGCLFPHRPGDSERI-EGARACVASHNRKALDSDSRFALLFG 124
QY 134 TDFNIDSLPFRKAAHMDVAVHESSPKANKYLFHKPVITLFFNYTATFSRSHLPLTTQYL 193
DB 125 TDFRASAAPLPLRLAHQSWALHDESSPLNNFLLSHGPGIRLFNLTSTFFSSHSDVPJSLQWL 184
```

```

QY 194 ESIEVLKSLRYLVLPQSKNKLKR-IAPLVYVQSDCDPSPSRDSYVELMTYTEVDSYGE 252
DB 185 PGTATLIR-RVYPPMEAEWRKRGYAPLVLQSHCHVDPADRDYVELNRHLPVDSYK 242
QY 253 CLRKNDLP-QQIKNPASMDADG--FYRIACYKFLAFENAVCDYITTEKWRPPLKLGIV 309
DB 243 CLNRELPTARLQDPAATATTEDPELLAFISRYKFLALENAICNDVYTELWPMHLGAV 302
QY 310 PYYGSPSITDMLPBNKSAIIYSEFSHPREIASIRLDSDDSLKLYEYVWKLKGLISNQ 369
DB 303 PYYRSPSPVRDMPWNHSLVILIDPESPOKLAEPIDFLDKNDEYMKYLAVKOPGGITNQ 362
QY 370 RLTLRERKRWGVDPVDNDYIDAFECNVCTKWANIRLQE-----KGLPP--KRWDAED 422
DB 363 FLDSLKREWGVDNPLLPNYLNGFEFCVDCYELARLADAKNAASPGDSFVEPHIAG 422
QY 423 THLSCEPFTVAFSPRLRPPLSLREWVTSFEOSKKAQALRWLYDRNCRSSQFFW 480
DB 423 SHMDCEVPTPGFNVVEIIPENDSWKEMLQDYWGQDGLQDGLTAMIHNET-EQTRFW 479
```

```

RESULT 7
US-10-467-595-7
; Sequence 7, Application US/10467595
; Publication No. US20040166501A1
; GENERAL INFORMATION:
; APPLICANT: AZIMZAI, Valda; YUE, Henry;
; APPLICANT: DING, Li; NGUYEN, Daniel B.;
; APPLICANT: GANDHI, Ameena R.; BURROD, Neil;
; APPLICANT: THANGAVELU, Kavitha; ELLIOTT, Vicki S.;
; APPLICANT: RAMKOMAR, Jayalaxmi; YAO, Monique G.;
; APPLICANT: LAU, Preeti G.; TANG, Y. Tom;
; APPLICANT: SWARNAKAR, Anita; WARREN, Bridget A.;
; APPLICANT: WALIA, Narinder K.; POLICKY, Jennifer L.;
; APPLICANT: XU, Yumeng; HONCHELL, Cynthia D.;
; APPLICANT: AU-YOUNG, Janice K.; BAUGHN, Mariah R.;
; APPLICANT: DUGGAN, Brendan M.; LU, Pyung Aina M.;
; APPLICANT: GIETZEN, Kimberly J.; JACKSON, Jennifer L.;
; APPLICANT: RAUWMANN, Bridget E.; LU, Yan;
; APPLICANT: KAREHT, Stephanie K.; TRAN, Uyen K.;
; APPLICANT: RICHARDSON, Thomas W.; EMBERTING, Brook M.;
; APPLICANT: HAPALIA April J.A.; BURRILL, John D.;
; APPLICANT: MARCUS, Gregory A.; ZINGLER, Kurt A.;
; APPLICANT: KABLE, Amy E.; CORVAD, Ann E.
; TITLE OF INVENTION: RECEPTORS AND MEMBRANE-ASSOCIATED PROTEINS
; FIDE REFERENCE: PF-0894 PCT
; CURRENT APPLICATION NUMBER: US/10/467,595
; PRIOR FILING DATE: 2003-08-06
; PRIOR APPLICATION NUMBER: PCT/US02/03868
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/267,201
; PRIOR FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: US 60/269,580
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/282,679
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/288,295
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/348,687
; PRIOR FILING DATE: 2002-01-14
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 1953366CD1
US-10-467-595-7
```

```

Query Match      21.1%; Score 554.5; DB 16; Length 168;
Best Local Similarity 76.8%; Pred. No. 1.1e-45;
```



Matches 106; Conservative 9; Mismatches 22; Indels 1; Gaps 1;

QY 34 EWMELGFEKFEKSSSSLODCHTMEAPTHLNSLKEGGLFNKRRKMDSPYIMLW 93  
DB 27 QVMELGFEKFEKSSSSLODCHTMEAPTHLNSFKEGLFNKRRKMDSPYIMLW 86  
QY 94 WSPLTGEGRLGCGACACFTTINRTYHHMTKAPLFYV-TDENDSLPLPRKADHWA 152  
DB 87 WSPLTGEGRLGCGACACFTTINRTYHHMTKAPLFYVLTQKNGRGARYPELPRAHS 146

QY 153 VFHEBSPNNYKTLHKPV 170  
DB 147 VCFLLTPDSTPELPARDV 164

RESULT 8  
US-09-764-877-1210  
Sequence 1210, Application US/09764877  
Patent No. US20020147140A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC005  
CURRENT APPLICATION NUMBER: US/09/764,877  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - refer to PALM or file wrapper  
NUMBER OF SEQ ID NOS: 4031  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1210  
LENGTH: 107  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-764-877-1210

Query Match

Best Local Similarity 19.8%; Score 521; DB 9; Length 107;  
Matches 100; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 178 ATFSHSHLPLTTOYLESIEVLKSLRYVPLQSKNKLRLAPLVYVQSDCPSPSRDSY 237  
DB 1 ATFSHSHLPLTTOYLESIEVLKSLRYVPLQSKNKLRLAPLVYVQSDCPSPSRDSY 60

QY 238 VRELMTYIEVDSYGECLRNKDLPOOLKNPASMADGFFYRII 278  
DB 61 VRELMTYIEVDSYGECLRNKDLPOOLKNPASMADGFFYRII 101

RESULT 9  
US-10-242-515-1210  
Sequence 1210, Application US/10242515  
Publication No. US20040009488A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC005C1  
CURRENT APPLICATION NUMBER: US/10/242,515  
CURRENT FILING DATE: 2002-09-13  
Prior application data removed - refer to PALM or file wrapper  
Prior application number: 09/764,877  
Prior filing date: 2001-01-17  
Prior application number: 60/179,065  
Prior filing date: 2000-01-31  
Prior application number: 60/180,628  
Prior filing date: 2000-02-04  
Prior application number: 60/214,886  
Prior filing date: 2000-06-28  
Prior application number: 60/217,487  
Prior filing date: 2000-07-11  
Prior application number: 60/225,758  
Prior filing date: 2000-08-14  
Prior application number: 60/220,963  
Prior filing date: 2000-07-26  
Prior application number: 60/217,496  
Prior filing date: 2000-07-11

Prior application number: 60/225,447  
Prior filing date: 2000-08-14  
Prior application number: 60/218,290  
Prior filing date: 2000-07-14  
Remaining prior application data removed - See file wrapper or PALM.  
Number of seq id nos: 4031  
Software: PatentIn Ver. 2.0  
Seq id no 1210  
Length: 107  
Type: PRT  
Organism: Homo sapiens  
US-10-242-515-1210

Query Match

Best Local Similarity 19.8%; Score 521; DB 15; Length 107;  
Matches 100; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 178 ATFSHSHLPLTTOYLESIEVLKSLRYVPLQSKNKLRLAPLVYVQSDCPSPSRDSY 237  
DB 1 ATFSHSHLPLTTOYLESIEVLKSLRYVPLQSKNKLRLAPLVYVQSDCPSPSRDSY 60

QY 238 VRELMTYIEVDSYGECLRNKDLPOOLKNPASMADGFFYRII 278  
DB 61 VRELMTYIEVDSYGECLRNKDLPOOLKNPASMADGFFYRII 101

RESULT 10  
US-10-072-977-21  
Sequence 21, Application US/10072977  
Publication No. US20030152933A1  
GENERAL INFORMATION:  
APPLICANT: Barash et al.  
TITLE OF INVENTION: Human Secreted Proteins  
FILE REFERENCE: PTO57P1  
CURRENT APPLICATION NUMBER: US/10/072,977  
CURRENT FILING DATE: 2002-02-12  
Prior application number: PCT/US01/25288  
Prior filing date: 2001-08-13  
Prior application number: US 60/225,215  
Prior filing date: 2000-08-14  
Number of seq id nos: 26  
Software: PatentIn Ver. 3.1  
Seq id no 21  
Length: 261  
Type: PRT  
Organism: Homo sapiens  
US-10-072-977-21

Query Match

Best Local Similarity 18.6%; Score 487.5; DB 14; Length 261;  
Matches 98; Conservative 51; Mismatches 89; Indels 11; Gaps 5;

QY 242 MTYIEVDSYGECLRNKDLPOOLKNPASMADG--FYRIIAYKFIAPFNAYCDDYITE 298  
DB 1 MRHLPVDSYGECLRNKDLPOOLKNPASMADG--FYRIIAYKFIAPFNAYCDDYITE 60

QY 299 KFWRLKGLVVPVYVGSPIIDWLPNSKSAIIVSEFHPRELASLYRRDSDLYEAYV 358  
DB 61 KLMRPMHIGAVPYVYVGSPIIDWLPNSKSAIIVSEFHPRELASLYRRDSDLYEAYV 120

QY 359 EYKLGELISNQLTALBERKGVQDVNQNIDAFECMCYCTKMAIRIOE-----KGL 413  
DB 121 AYKOPGGINQFLDLSLGRWGVNDPLLPYVNLGFCFCDYELARLDKKAHAASPGD 180

QY 414 P--KRWAEADTHLSCPPYFAFSPRTPLSLRSMWISSFEQSKKEAQAALRWLVDRN 471  
DB 181 SVFEPHIAQSHNDQVPTPGFVAVBEIPENDSKEMWIODYWGQDQGEALTAHINN 240

QY 472 QNFSSQEFW 480  
DB 241 ET-EQTKFW 248

```
RESULT 11
US-10-080-960-17
; Sequence 17, Application US/10080960
; Publication No. US20020197695A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Gluckmann, Maria
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 80090, 52874, 52880, 63497, AND 33425
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES THEREOF
; FILE REFERENCE: 38155-20044.00
; CURRENT APPLICATION NUMBER: US/10/080,960
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/242,040
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/242,038
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/241,992
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/242,637
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus amino acid sequence
US-10-080-960-17

Query Match      14.1%; Score 369; DB 13; Length 178;
Best Local Similarity 42.4%; Pred. No. 1.7e-27;
Matches 75; Conservative 31; Mismatches 67; Indels 4; Gaps 3;

QY 91 MLMWS-PLTGETGRJLQCCADACFTTINKTYLHMHMTAFPLPYGTDENIDSLPRCAHH 149
DB 4 LVMWSRDMGMNYDVORQCGINTCRITNKRS--RRPWAGVLFYSGNKTGDPFLPRVNEHQ 61
QY 150 DNAFHEESPKNKYLFHKPVITLTFNYTATFSRSHSLPLTQYTESIEVLSKRLYVPLQ 209
DB 62 IVALHHEESPNTPEFSNKEFLRHHFTSTSRYSNLPITMYLPSEGLTSKDYVTFD 121
QY 210 SKKLRKR-LAPLVYVQSDCCPPSRDSYVRELMTYIEVDSYGECLRNKDLPOQLN 265
DB 122 GSKXGVRPSTSVFLQSDCDTMSGREDEVYKELMKHLPIDSYGLSRNRLPEROKO 178

RESULT 12
US-10-108-260A-3173
; Sequence 3173, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cdna
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3173
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3173

Query Match      12.1%; Score 318.5; DB 15; Length 157;
Best Local Similarity 43.4%; Pred. No. 1.3e-22;
Matches 62; Conservative 26; Mismatches 48; Indels 7; Gaps 2;

QY 296 ITEXWRPLKGVVYVYGGPSITDMLPSNKSALVSEFSPREBLASVYRRLDSDRLYE 355
DB 1 MTEKLRPMHLGAIVYVYGGPSVADWMFNHNSVILLIDFESPOKLAEIFIDFLKNDDEYM 60
```

```
QY 356 AVEWKLGEISNQLLTALREKRGVQDVQNDYIDAFECMVCTKWANIRLOE----- 410
DB 61 KYLAAYKQPGGINTNGLDLSLKHREGVNDPLPRYVINGECFVCDYELALDEKXHAAS 120
QY 411 KGLRP--KWEAEQTHLSCEPPT 431
DB 121 PCDSPVEFHIAQPSHMDCPVPT 143

RESULT 13
US-10-120-319-9
; Sequence 9, Application US/10120319
; Publication No. US20020164749A1
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/10/120,319
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/092,315
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Bos taurus
US-10-120-319-9

Query Match      10.8%; Score 282.5; DB 13; Length 365;
Best Local Similarity 28.4%; Pred. No. 1.5e-18;
Matches 97; Conservative 49; Mismatches 117; Indels 79; Gaps 16;

QY 90 IMLWSPITGETGRJLQCC-----GADACFTINRT-----YLHH-----HNTKAFLY 132
DB 69 VLMWTWPR-NQPVALSCKSELMLWGTQCLTVRSEYPOADAVFVHHREVSRRPQQL-- 125
QY 133 GTDENIDSLPRCAHHDMAVFHEESPKNKYLFHKPVITLTFNYTATFSRSHSLPLTQY 192
DB 126 -----PSPRPADRWVWFMSFSPNCKLK--KDLGYFNLITWYRSDSIFWPYGM 175
QY 193 LESIEVLSKRLYVPLQSKKRLRLAPLVYVQSDCCPPSRDSYVRELMTYIEVDSYGE 252
DB 176 LEPMP-SQPVETILNISKAKTL-----VANVSNMNTDSIRVQYKILKPHLQVDVYGR 228
QY 253 CLRNKDLPOQLKNPASMDADGFYRIIAQYFYILAFENAVCDYITEKFWR-PLKLGVPV 311
DB 229 F--HTPLPHAL-----MAKQLSQYFYFLAFENSLHPDYITKLMKNLQAWAVPV 276
QY 312 YVGSPEST--DWLPNSKALVSEFSPREBLASVYRRLDSDRLYEAVYEMKLGEISNQ 369
DB 277 VLGSRVNYQGLRP-KAFIHVEDFQSPKDLAQLLADKDYASVINYFRWR----- 327
QY 370 RLTLALREKRGVQDVQNDYIDAFECMVCTKWANIRLOEK 411
DB 328 ---ETLRPSF-----SWALMCKXCM--KLQGE 351

RESULT 14
US-10-189-977-9
; Sequence 9, Application US/10189977
; Publication No. US20030166211A1
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/10/189,977
; CURRENT FILING DATE: 2002-07-03
```

PRIOR APPLICATION NUMBER: US/09/092,315  
 PRIOR FILING DATE: 1998-06-05  
 PRIOR APPLICATION NUMBER: US 60/048,857  
 PRIOR FILING DATE: 1997-06-06  
 NUMBER OF SEQ ID NOS: 22  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 9

LENGTH: 365

TYPE: PRT

ORGANISM: Bos taurus

US-10-189-977-9

Query Match 10.8%; Score 282.5; DB 14; Length 365;

Best Local Similarity 28.4%; Pred. No. 1.5e-18;  
 Matches 97; Conservative 49; Mismatches 117; Indels 79; Gaps 16;

90 IMWMSPLTGTGTLGQC-----GADACFTTNR-----YLH-----HMTKAFLEY 132  
 69 VLLMTWPF-NQPVALSRCSELMPTGADCOLTVNRSEYPOADAVFVHREVSHPKQJL-- 125  
 133 GTDFNIDSLPLPRKAHDMWAVFHEESPKNKYKLFHKPVITLFPNYATFSPRSHLPLTTQY 192  
 126 -----PPSPRPADQGRWVFMESPNCGLK--KLDGYNFLNLSYRSDSDIFMPYGM 175  
 193 LESIEVLKSLRYLVPLOSKNKLRLAPLVYVQSDCDPSPDSYVRELMTYIEVDSIGE 252  
 176 LEPWP-SQPVETLLNISAKTKL-----VAVVSNWNTDSIRVQYKLLKPHLOVDVYGR 228  
 253 CLRNKDLPOOLKNPASMDDGFYRIIAQYKFLAFENAVCDYITEKFWR-PLKLGVPV 311  
 229 F--HTPLPHAL-----MAKQLSQKFLAFENSLHPDYITEKMKNALQAMAVPV 276  
 312 YGSPSIT--DWLPNSKSAIIVSEFSPRELASYIRLDSDDLRYEAYVEWKLGEISNQ 369  
 277 VLGSRVNYEQFLPP-KAFIHVEDFQSPKDLAQYLALDKDYASYLNYFRWR----- 327  
 370 RLTLALRERKMGVQDVQNDYIDAFCMCVCTKVMANIRLOEK 411  
 328 ---ETLRFRSF-----SWALMFCKACW--KLQOE 351

RESULT 15

US-10-392-098-9

Sequence 9, Application US/10392098

Publication No. US20030166212A1

GENERAL INFORMATION:

APPLICANT: Taylor, Diane E.

APPLICANT: Ge, Zhongming

TITLE OF INVENTION: NUCLEIC ACIDS ENCODING ALPHA-1.3

TITLE OF INVENTION: FUCOSYLTRANSFERASES AND EXPRESSION SYSTEMS FOR MAKING AND

TITLE OF INVENTION: EXPRESSING THEM (amended)

FILE REFERENCE: 07254-049002

CURRENT APPLICATION NUMBER: US/10/392,098

CURRENT FILING DATE: 2003-03-17

PRIOR APPLICATION NUMBER: US/09/733,524A

PRIOR FILING DATE: 2000-12-07

PRIOR APPLICATION NUMBER: US 09/092,315

PRIOR FILING DATE: 1998-06-05

PRIOR APPLICATION NUMBER: US 60/048,857

PRIOR FILING DATE: 1997-06-06

NUMBER OF SEQ ID NOS: 27

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 9

LENGTH: 365

TYPE: PRT

ORGANISM: Bos taurus

US-10-392-098-9

Query Match 10.6%; Score 277.5; DB 14; Length 365;

Best Local Similarity 28.4%; Pred. No. 4.6e-18;  
 Matches 97; Conservative 48; Mismatches 118; Indels 79; Gaps 16;

90 IMWMSPLTGTGTLGQC-----GADACFTTNR-----YLH-----HMTKAFLEY 132

69 VLLMTWPF-NQPVALSRCSELMPTGADCOLTVNRSEYPOADAVFVHREVSHPKQJL-- 125  
 133 GTDFNIDSLPLPRKAHDMWAVFHEESPKNKYKLFHKPVITLFPNYATFSPRSHLPLTTQY 192  
 126 -----PPSPRPADQGRWVFMESPNCGLK--KLDGYNFLNLSYRSDSDIFMPYGM 175  
 193 LESIEVLKSLRYLVPLOSKNKLRLAPLVYVQSDCDPSPDSYVRELMTYIEVDSIGE 252  
 176 LEPWP-SQPVETLLNISAKTKL-----VAVVSNWNTDSIRVQYKLLKPHLOVDVYGR 228  
 253 CLRNKDLPOOLKNPASMDDGFYRIIAQYKFLAFENAVCDYITEKFWR-PLKLGVPV 311  
 229 F--HTPLPHAL-----MAKQLSQKFLAFENSLHPDYITEKMKNALQAMAVPV 276  
 312 YGSPSIT--DWLPNSKSAIIVSEFSPRELASYIRLDSDDLRYEAYVEWKLGEISNQ 369  
 277 VLGSRVNYEQFLPP-KAFIHVEDFQSPKDLAQYLALDKDYASYLNYFRWR----- 327  
 370 RLTLALRERKMGVQDVQNDYIDAFCMCVCTKVMANIRLOEK 411  
 328 ---ETLRFRSF-----SWALMFCKACW--KLQOE 351

Search completed: September 9, 2004, 21:58:05  
 Job time : 132 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: September 9, 2004, 21:41:56 ; Search time 34 Seconds

(without alignments)  
737,948 Million cell updates/sec

Title: US-10-080-960-2

Perfect score: 2625

Sequence: 1 MKYTGPPQGVTDNMQCFNDQ.....LYDRNQNPFSSQFNGVLVRKD 486

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A COMB pep:\*

2: /cgn2\_6/ptodata/2/1aa/5B COMB pep:\*

3: /cgn2\_6/ptodata/2/1aa/6A COMB pep:\*

4: /cgn2\_6/ptodata/2/1aa/6B COMB pep:\*

5: /cgn2\_6/ptodata/2/1aa/PTUG COMB pep:\*

6: /cgn2\_6/ptodata/2/1aa/backfile1 pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description       |
|------------|-------|-------------|--------|----|-------------------|
| 1          | 282.5 | 10.8        | 365    | 4  | US-09-092-315-9   |
| 2          | 282.5 | 10.8        | 365    | 4  | US-09-390-131-7   |
| 3          | 277.5 | 10.6        | 365    | 4  | US-09-733-524A-9  |
| 4          | 263   | 10.0        | 359    | 1  | US-07-914-281-14  |
| 5          | 263   | 10.0        | 359    | 1  | US-08-393-246-14  |
| 6          | 263   | 10.0        | 359    | 1  | US-08-525-058A-14 |
| 7          | 263   | 10.0        | 359    | 2  | US-08-696-731-14  |
| 8          | 263   | 10.0        | 359    | 3  | US-09-042-531-14  |
| 9          | 263   | 10.0        | 359    | 4  | US-09-092-315-10  |
| 10         | 263   | 10.0        | 359    | 4  | US-09-733-524A-10 |
| 11         | 255.5 | 9.7         | 450    | 4  | US-09-390-131-8   |
| 12         | 255.5 | 9.7         | 450    | 4  | US-09-390-131-9   |
| 13         | 255.5 | 9.7         | 451    | 4  | US-09-390-131-3   |
| 14         | 251   | 9.6         | 356    | 4  | US-09-092-315-12  |
| 15         | 245.5 | 9.4         | 374    | 1  | US-07-914-281-11  |
| 16         | 245.5 | 9.4         | 374    | 1  | US-08-393-246-11  |
| 17         | 245.5 | 9.4         | 374    | 2  | US-08-525-058A-11 |
| 18         | 245.5 | 9.4         | 374    | 2  | US-08-696-731-11  |
| 19         | 245.5 | 9.4         | 374    | 2  | US-09-042-531-11  |
| 20         | 239.5 | 9.1         | 299    | 5  | PCT-US91-00899-6  |
| 21         | 239.5 | 9.1         | 361    | 1  | US-07-914-281-2   |
| 22         | 239.5 | 9.1         | 361    | 1  | US-08-393-246-2   |
| 23         | 239.5 | 9.1         | 361    | 1  | US-08-733-524-3   |
| 24         | 239.5 | 9.1         | 361    | 1  | US-08-525-058A-2  |
| 25         | 239.5 | 9.1         | 361    | 2  | US-08-696-731-2   |
| 26         | 239.5 | 9.1         | 361    | 3  | US-09-042-531-2   |
| 27         | 239.5 | 9.1         | 361    | 4  | US-09-390-131-6   |

|    |       |     |     |   |                   |                   |
|----|-------|-----|-----|---|-------------------|-------------------|
| 28 | 239.5 | 9.1 | 361 | 5 | PCT-US91-00899-7  | Sequence 7, Appl  |
| 29 | 237.5 | 9.0 | 355 | 4 | US-09-733-524A-12 | Sequence 12, Appl |
| 30 | 233.5 | 8.9 | 342 | 2 | US-08-483-151-2   | Sequence 2, Appl  |
| 31 | 219   | 8.3 | 405 | 1 | US-07-914-281-8   | Sequence 8, Appl  |
| 32 | 219   | 8.3 | 405 | 1 | US-08-393-246-8   | Sequence 8, Appl  |
| 33 | 219   | 8.3 | 405 | 1 | US-08-525-058A-8  | Sequence 8, Appl  |
| 34 | 219   | 8.3 | 405 | 2 | US-08-696-731-8   | Sequence 8, Appl  |
| 35 | 219   | 8.3 | 405 | 2 | US-09-042-531-8   | Sequence 8, Appl  |
| 36 | 215   | 8.2 | 405 | 2 | US-08-483-151-4   | Sequence 8, Appl  |
| 37 | 213   | 8.1 | 433 | 4 | US-09-092-315-11  | Sequence 11, Appl |
| 38 | 213   | 8.1 | 433 | 4 | US-09-733-524A-11 | Sequence 11, Appl |
| 39 | 197.5 | 7.5 | 357 | 5 | PCT-US91-00899-14 | Sequence 14, Appl |
| 40 | 150   | 5.7 | 440 | 4 | US-09-392-315-3   | Sequence 3, Appl  |
| 41 | 150   | 5.7 | 440 | 4 | US-09-733-524A-3  | Sequence 3, Appl  |
| 42 | 143   | 5.4 | 425 | 4 | US-09-092-315-6   | Sequence 6, Appl  |
| 43 | 143   | 5.4 | 425 | 4 | US-09-733-524A-6  | Sequence 6, Appl  |
| 44 | 142   | 5.4 | 476 | 4 | US-09-092-315-5   | Sequence 5, Appl  |
| 45 | 142   | 5.4 | 476 | 4 | US-09-733-524A-5  | Sequence 5, Appl  |

#### ALIGNMENTS

```
RESULT 1
US-09-092-315-9
; Sequence 9, Application US/09092315
; Patent No. 6399337
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/09/092,315
; CURRENT FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: US 60/048,857
; EARLIER FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-092-315-9
```

```
Query Match 10.8%; Score 282.5; DB 4; Length 365;
Best Local Similarity 28.4%; Pred. No. 6,8e-21;
Matches 97; Conservative 49; Mismatches 117; Indels 79; Gaps 16;

QY 90 IMWMSPLTGTGRLGOC-----GADACFTINRT-----YLHH-----HMTKAFLEFY 132
DB 69 VLLMTWPF-NQPVVLSKSELMFGTADCOITVRSEYQADAVFVHHRVSHRPKQL-- 125
QY 133 GTDFNIDSLPRKRAHDMAVFHESEPRNNYKLFHKPVITLFNVTATFSRSHLPLTTQY 192
DB 126 -----PSPRPADQQRWTFWSESPSNCILK--KLDGYFNILTWYSRSDIFMPYGW 175
QY 193 LESIEVLKSLRYVPLDQSKTKRLKRLPIVYVOSDQPPSDPSYVREMTYIEVSIGE 252
DB 176 LEPWP-SQPVETLLNISAKTKL-----VAMVVSNMWTDISIRQYVYLLPHQVVDYGR 228
QY 253 CLRNKDLPOQKNPASMADGFGYRIIAQYKFIILAFENAVCDYITEKFWP-PLKGVVPV 311
DB 229 F-HHPLHAL-----MAKQIQYKRYLAFENSLHPDYITETKLMKNAQAMAVPV 276
QY 312 YYGSPSIT-DWLPENKSAIIVSESPHREPLASTIRLSDSDRIYKZYVWVKLKGELSNQ 369
DB 277 VLGPRVYVEQFLP-KAFIHEDEFQSKDLAQYLLDLDDYASLYNVPFWR----- 327
QY 370 RLTLALREKMGQVQVQNDNYIDAFECYVCYKVNANIRLOEK 411
DB 328 ---ETLRPSF-----SWALMFCAKCV---KLQGE 351
```

Mon Sep 13 09:12:38 2004

us-10-080-960-2.ra1

Page 2

```
RESULT 2
US-09-390-131-7
; Sequence 7, Application US/09390131
; Patent No. 6461835
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Cummings, Richard D.
; APPLICANT: Nyame, A. Kwame
; APPLICANT: DeRose-Boyd, Russell A.
; TITLE OF INVENTION: FUCOSYLTRANSFERASES, POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING FUCOSYLTRANSFERASES, AND TRANSGENIC MAMMAL
; TITLE OF INVENTION: INCORPORATING SAME
; FILE REFERENCE: 6679.US.01
; CURRENT APPLICATION NUMBER: US/09/390,131
; CURRENT FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-390-131-7

Query Match      10.8%; Score 282.5; DB 4; Length 365;
Best Local Similarity 28.4%; Pred. No. 6.8e-21;
Matches 97; Conservative 49; Mismatches 117; Indels 79; Gaps 16;

QY 90 INLWMSPLTGETGRGQC-----GADACFTTINT-----YLHH-----HMTKAFLFY 132
DB 69 VILMTWPF-NQPVALSRCSELMPTGADCOLTVNRESEYQADAVFVHREVSHPKQCL-- 125
QY 133 GTDFNIDSLPLPRKXAHMDAVFHESESPKNNYKLFHKPVITLFNTATFSRSHLPLTTOY 192
DB 126 -----PSPRPADQKWVWFMSWSPNSCKL--KLDGYFNLTSTRSDIFMPYGM 175
QY 193 LESIVLKSILYVLPLOSNNKLRKRLAPLVYVQSDCPSPDRDSYVREIMTYIEVDSYGE 252
DB 176 LEPWP-SQVETLNLISAKTKL-----VAMVSNMNTDSIRVOYKTLKPHLQVDVYGR 228
QY 253 CLRNKDLPOQJKNPASMADGFRIRIAQYKFIILAFENAVCCDYITEKFWR-PLKLGVPVY 311
DB 229 F--HTPLPHAL-----MAKQLSQYKFIILAFENSLHPDYITTEKLMKNALQAMAVPV 276
QY 312 YYGSPSIT--DWLPNKSALIVSEFHSHPRELASTYIRLSDSDRLYEAYVEWKLGELISNQ 369
DB 277 VLGPSRVNVEQFLPP-KAFIHVEDFQSPKDLAQYLLALDKDYASYLYNFRWR----- 327
QY 370 RLTLALRERKKGVDVQNDVNIADAFECMVCTKWMANIRLOEK 411
DB 328 ---ETLRPRSF-----SWALMFCKACW---KLQOE 351

RESULT 3
US-09-733-524A-9
; Sequence 9, Application US/09733524A
; Patent No. 6534298
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING ALPHA-1.3
; TITLE OF INVENTION: FUCOSYLTRANSFERASES AND EXPRESSION SYSTEMS FOR MAKING AND
; TITLE OF INVENTION: EXPRESSING THEM (amended)
; FILE REFERENCE: 07254-049002
; CURRENT APPLICATION NUMBER: US/09/733,524A
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 09/092,315
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 365
```

```
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-733-524A-9

Query Match      10.6%; Score 277.5; DB 4; Length 365;
Best Local Similarity 28.4%; Pred. No. 2.3e-20;
Matches 97; Conservative 48; Mismatches 118; Indels 79; Gaps 16;

QY 90 INLWMSPLTGETGRGQC-----GADACFTTINT-----YLHH-----HMTKAFLFY 132
DB 69 VILMTWPF-NQPVALSRCSELMPTGADCOLTVNRESEYQADAVFVHREVSHPKQCL-- 125
QY 133 GTDFNIDSLPLPRKXAHMDAVFHESESPKNNYKLFHKPVITLFNTATFSRSHLPLTTOY 192
DB 126 -----PSPRPADQKWVWFMSWSPNSCKL--KLDGYFNLTSTRSDIFMPYGM 175
QY 193 LESIVLKSILYVLPLOSNNKLRKRLAPLVYVQSDCPSPDRDSYVREIMTYIEVDSYGE 252
DB 176 LEPWP-SQVETLNLISAKTKL-----VAMVSNMNTDSIRVOYKTLKPHLQVDVYGR 228
QY 253 CLRNKDLPOQJKNPASMADGFRIRIAQYKFIILAFENAVCCDYITEKFWR-PLKLGVPVY 311
DB 229 F--HTPLPHAL-----MAKQLSQYKFIILAFENSLHPDYITTEKLMKNALQAMAVPV 276
QY 312 YYGSPSIT--DWLPNKSALIVSEFHSHPRELASTYIRLSDSDRLYEAYVEWKLGELISNQ 369
DB 277 VLGPSRVNVEQFLPP-KAFIHVEDFQSPKDLAQYLLALDKDYASYLYNFRWR----- 327
QY 370 RLTLALRERKKGVDVQNDVNIADAFECMVCTKWMANIRLOEK 411
DB 328 ---ETLRPRSF-----SWALMFCKACW---KLQOE 351

RESULT 4
US-07-914-281-14
; Sequence 14, Application US/07914281
; Patent No. 5324663
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, WAIER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/914,281
; FILING DATE: 19920720
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavelle, Jean-Paul W. P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-060-55
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELE: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: AMINO ACID
```

TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-07-914-281-14

Query Match 10.0%; Score 263; DB 1; Length 359;  
Best local Similarity 27.4%; Pred. No. 7.4e-19;  
Matches 98; Conservative 48; Mismatches 118; Indels 94; Gaps 19;

87 SYPIMLWSPLTGRTGLGQC-----GADACFEFTNR-----TYLHHMTKAFLEFG 133  
60 SIPILMTWPFNKPIALPRCSEWPGTADCNITADRKYPQADAVIYHRE-----VMYN 115  
134 TDFNIDSLP-LPRKAHDMVAFHEESPKNKYLFHKVITLTENYTFTRSHLPITTY 192  
116 PS--AQLPSRPRQGQRFWTFMESPSHCWOL--KAMGYFN-----LTMSY 158  
193 LESIEVLKSLRYLP-----LOSKNKLRKILAPLVYVQSDCDPPSRDSYREL 241  
159 RSDSDITFTPGWLEPWSGQPAHPPLNLSAKTEL-----VANAWSWGNSARVRYOQL 212  
242 MTYIEVDSYGECLENNKDLPOQLKNPASMADGFRITIAQYKFLAENAVCDYITEKFW 301  
213 QAHKLVVYGR--SHKPLPQ-----GTWMETLSRYKFLAENSLHPDYITEKLM 260  
302 R-PLKLGVPVYVYGSPSITDW---LPSNKSALIVSEFSHPRELASIYIRLSDSDRLYEAY 357  
261 RNALFAMVAVPVVLG-PSRSNYERFLPD-AFIHVDDFQSKDLARYLOELDKDHARYLSY 318  
358 VEWLKGEISNORLLTLALRER--KMGVQDVNQNDYIDAFECMWCTKWANIRLOEKGL 413  
319 FRWR-----ETLRPRFSFMAL-----AF-CKACWKLQESRYQTRGI 354

RESULT 5  
US-08-393-246-14  
Sequence 14, Application US/08393246  
Patent No. 5595900

GENERAL INFORMATION:  
APPLICANT: LOWE, JOHN B.  
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS  
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,  
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION  
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCT  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 Jefferson Davis Highway, Fourth floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/393,246  
FILING DATE:  
CLASSIFICATION: 510  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/220,433  
FILING DATE: 30-MAR-1994  
APPLICATION NUMBER: US 07/914,281  
FILING DATE: 20-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Lavalleye, Jean-Paul M. P.  
REGISTRATION NUMBER: 31,451  
REFERENCE/DOCKET NUMBER: 2363-060-55  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 521-4500  
TELEFAX: (703) 486-2347

TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 359 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-393-246-14

Query Match 10.0%; Score 263; DB 1; Length 359;  
Best local Similarity 27.4%; Pred. No. 7.4e-19;  
Matches 98; Conservative 48; Mismatches 118; Indels 94; Gaps 19;

87 SYPIMLWSPLTGRTGLGQC-----GADACFEFTNR-----TYLHHMTKAFLEFG 133  
60 SIPILMTWPFNKPIALPRCSEWPGTADCNITADRKYPQADAVIYHRE-----VMYN 115  
134 TDFNIDSLP-LPRKAHDMVAFHEESPKNKYLFHKVITLTENYTFTRSHLPITTY 192  
116 PS--AQLPSRPRQGQRFWTFMESPSHCWOL--KAMGYFN-----LTMSY 158  
193 LESIEVLKSLRYLP-----LOSKNKLRKILAPLVYVQSDCDPPSRDSYREL 241  
159 RSDSDITFTPGWLEPWSGQPAHPPLNLSAKTEL-----VANAWSWGNSARVRYOQL 212  
242 MTYIEVDSYGECLENNKDLPOQLKNPASMADGFRITIAQYKFLAENAVCDYITEKFW 301  
213 QAHKLVVYGR--SHKPLPQ-----GTWMETLSRYKFLAENSLHPDYITEKLM 260  
302 R-PLKLGVPVYVYGSPSITDW---LPSNKSALIVSEFSHPRELASIYIRLSDSDRLYEAY 357  
261 RNALFAMVAVPVVLG-PSRSNYERFLPD-AFIHVDDFQSKDLARYLOELDKDHARYLSY 318  
358 VEWLKGEISNORLLTLALRER--KMGVQDVNQNDYIDAFECMWCTKWANIRLOEKGL 413  
319 FRWR-----ETLRPRFSFMAL-----AF-CKACWKLQESRYQTRGI 354

RESULT 6  
US-08-525-058A-14  
Sequence 14, Application US/08525058A  
Patent No. 5770420

GENERAL INFORMATION:  
APPLICANT: LOWE, JOHN B.  
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS  
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,  
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION  
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCT  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.  
STREET: 1755 Jefferson Davis Highway, Fourth floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/525,058A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lavalleye, Jean-Paul M. P.  
REGISTRATION NUMBER: 31,451  
REFERENCE/DOCKET NUMBER: 2363-060-55  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 521-4500  
TELEFAX: (703) 486-2347  
TELEX: 248855 OPAT UR

; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 359 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-525-058A-14

Query Match 10.0%; Score 263; DB 1; Length 359;  
 Best Local Similarity 27.4%; Pred. No. 7, 4e-19;  
 Matches 98; Conservative 48; Mismatches 118; Indels 94; Gaps 19;

QY 87 SYPIIMWSPITGETGRIGOC-----GADACFTINR-----TYLHMTKAFLEFG 133  
 Db 60 SIPILMTWPNKXIALPSCSEWVPGTADCNITADKRYVPOADAVIYHRE-----VMYN 115  
 QY 134 TDFNIDSLP-LPRKAHDMAVFHEESPKNKYLFHKFVITLNTATSRSHSLPLTQY 192  
 Db 116 PS---AQLPSRPRGQRMWFSWSPSHCWOL--KAMDGYFN-----LTWSY 158  
 QY 193 LESIEVLKSLRYLP-----LOSKNKLRKRLAPLVYVQSDCDPPSDRDSYVREL 241  
 Db 159 RSDSDIFTPYGMLEPMSGQPAHPPLNLSAKTEL-----VAVASWNGPNSARVRYQSL 212  
 QY 242 MTYIEVDSYGECLRNKDLPOQLKNPASMADGFYRIIAQYKFLAFENAVCDYITTEKFW 301  
 Db 213 QAHKLVQVYGR--SHKPLPQ-----GTMMETLSRYKFLAFENSLHPDYITEXLW 260  
 QY 302 R-PLKLGVPVYVYSPSITDW---LPSNKSAILVSESHPRELASYIRLSDRIYEXAY 357  
 Db 261 RNALBAMAVPVVLG-PSRSNYERFLPPD-AFIHVDDQSPKDLARYLOELDKHARYLSY 318  
 QY 358 VEWKLGKHSNQRLTLALRER--KMGVQVQNDNYIDAFECWCTFWMANIRLOEKGL 413  
 Db 319 FRWR-----ETLRPRSPFSWAL-----AF-CXACWKLQESRSRYQIRGI 354

## RESULT 7

US-08-696-731-14

; Sequence 14, Application US/08696731  
 ; Patent No. 595347

; GENERAL INFORMATION:  
 ; APPLICANT: LOWE, JOHN B.

; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,

; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OHLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

; STREET: 1755 Jefferson Davis Highway, Fourth Floor

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/696,731

; FILING DATE: 14-AUG-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/393,246

; FILING DATE: US 08/220,433

; APPLICATION NUMBER: 30-MAR-1994

; FILING DATE: US 07/914,281

; FILING DATE: 20-JUL-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Lavalleye, Jean-Paul M. P.  
 ; REGISTRATION NUMBER: 31,451  
 ; REFERENCE/DOCKET NUMBER: 2363-060-55  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703)521-4500  
 ; TELEFAX: (703)486-2347  
 ; TELEX: 248855 OPAT UR  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 359 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-696-731-14

Query Match 10.0%; Score 263; DB 2; Length 359;  
 Best Local Similarity 27.4%; Pred. No. 7, 4e-19;  
 Matches 98; Conservative 48; Mismatches 118; Indels 94; Gaps 19;

QY 87 SYPIIMWSPITGETGRIGOC-----GADACFTINR-----TYLHMTKAFLEFG 133  
 Db 60 SIPILMTWPNKXIALPSCSEWVPGTADCNITADKRYVPOADAVIYHRE-----VMYN 115  
 QY 134 TDFNIDSLP-LPRKAHDMAVFHEESPKNKYLFHKFVITLNTATSRSHSLPLTQY 192  
 Db 116 PS---AQLPSRPRGQRMWFSWSPSHCWOL--KAMDGYFN-----LTWSY 158  
 QY 193 LESIEVLKSLRYLP-----LOSKNKLRKRLAPLVYVQSDCDPPSDRDSYVREL 241  
 Db 159 RSDSDIFTPYGMLEPMSGQPAHPPLNLSAKTEL-----VAVASWNGPNSARVRYQSL 212  
 QY 242 MTYIEVDSYGECLRNKDLPOQLKNPASMADGFYRIIAQYKFLAFENAVCDYITTEKFW 301  
 Db 213 QAHKLVQVYGR--SHKPLPQ-----GTMMETLSRYKFLAFENSLHPDYITEXLW 260  
 QY 302 R-PLKLGVPVYVYSPSITDW---LPSNKSAILVSESHPRELASYIRLSDRIYEXAY 357  
 Db 261 RNALBAMAVPVVLG-PSRSNYERFLPPD-AFIHVDDQSPKDLARYLOELDKHARYLSY 318  
 QY 358 VEWKLGKHSNQRLTLALRER--KMGVQVQNDNYIDAFECWCTFWMANIRLOEKGL 413  
 Db 319 FRWR-----ETLRPRSPFSWAL-----AF-CXACWKLQESRSRYQIRGI 354

## RESULT 8

US-09-042-531-14

; Sequence 14, Application US/09042531  
 ; Patent No. 6268193

; GENERAL INFORMATION:  
 ; APPLICANT: LOWE, JOHN B.

; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,

; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OHLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

; STREET: 1755 Jefferson Davis Highway, Fourth Floor

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/042,531

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:



QY 833 TTGACAGTATTAAGTTATTCCTAGCTTTTGAGATGACGTTGTGATGCTACATCACTG 892  
 DB 361 TGGACAGTATTAAGTTATTCCTAGCTTTTGAGATGACGTTGTGATGCTACATCACTG 420  
 QY 893 AGAAGTTCTGGAGGCGCATGAAAAGTGGGGTAGTCCCTGATATTAATTCAGATCCCGCAGCA 952  
 DB 421 AAAAAGCTGGCGGCGCTGTGTGTGGAGTGGTACCTGTGTGCTTTGGTTCCCGCAGCA 480  
 QY 993 TCACAGACTGGCTTCGCAAGTAAAGTCTATTCTGTATGAGATTTTCTCACCCCA 1012  
 DB 481 TTATCGACTGGCTTCGCAAGTAAAGTCTATTCTGTATGAGATTTTCTCACCCCTC 540  
 QY 1013 GGAAGTGGGCAAGTATACAGAGAGCTGATCTGATGACAGATGATGAGGCTATG 1072  
 DB 541 GAGAGCTGGCGGCTTATATCAAAAGCTGATCAAAATGACGAGATGAGGCTTAC 600  
 QY 1073 TAGAATGGAAGCTGAA-GGGTGAATCTCTAACCGAGGACTTTGACAGCTCTCAGGGAA 1131  
 DB 601 TGGATGGAAGCTGAAAGAGGAGCATTTCCATCCAGGCTGCTTCACCAATGAGGAA 660  
 QY 1132 CGGAATGGGAGTGAAGAGAGCTCAACAGAGACAT-TACATGATGATGAGGCTAT 1190  
 DB 661 CGCAATGGGAGTGAAGAGATGCTCAAGAGATCTATGACACATTTGAGTGCT 720  
 QY 1191 GGATGCA 1198  
 DB 721 GGATGTA 728

## RESULT 15

EX873167 839 bp mRNA linear EST 17-DEC-2003

DEFINITION BX873167 AGENAE Rainbow trout multi-tissues-normalized (tcbk)

ACCESSION BX873167 Oncorhynchus mykiss cDNA clone tck0017c.p.23 5prim, mRNA sequence.

VERSION BX873167.1 GI:40001712

KEYWORDS EST.

SOURCE Oncorhynchus mykiss (rainbow trout)

ORGANISM Oncorhynchus mykiss

REFERENCE 1 (bases 1 to 839) Governor, M., Guiguen, Y. and Le Gac, F.

AUTHORS Construction and primary characterization of normalized cDNA

TITLE Libraries in rainbow trout, Oncorhynchus mykiss

JOURNAL Unpublished (2003)

COMMENT Contact: Guiguen Y

INRA - SCRIBE

Campus de beaulieu, RENNES cedex, 35042, France

Tel: 02.23.48.50.09

Fax: 02.23.48.50.20

Email: Yann.Guiguen@beaulieu.rennes.inra.fr

Sequence cleaned of vector, adaptor and repetitions. Contact us

at sigena@port@jouy.inra.fr to obtain the chromatogram of this

sequence.

Plate: 0017 row: p column: 23

Seq primer: M13R.

Location/Qualifiers

1. .839

/organism="Oncorhynchus mykiss"

/mol\_type="mRNA"

/db\_xref="taxon:8022"

/clone="tck0017c.p.23"

/tissue\_type="multi-tissues"

/dev\_stage="from embryos to adults"

/lab\_host="DH10B"

/clone\_id="AGENAE Rainbow trout multi-tissues-normalized (tcbk)"

/note="Vector: pT73D-pac; Rainbow trout multi-tissues-normalized + 2 subtractions; Clone distribution: AGENAE Resource centre, Francois PIUMI, Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et

## ORIGIN

Query Match

Best Local Similarity

Marches 534; Conservative

0; Mismatches 285; Indels 3; Gaps 1;

Etude du genome (JREG). Domaine de Vilvert, 78352, Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33 (0) 1.34.65.22.73"

QY 570 CCATATCTGGAGAGGATTAAGTCTGAAAGTCACTCCGATTCCTAGTTCCTTGCAGTC 629  
 DB 9 CCAGACCTGGAGAACTGAGAGCTCTGACCTCCAGACCTACCTGCTGCTCCCAACA 68  
 QY 630 CAAAAAGAGCTTGAAGAAAGCTTCTCGGCTGGTATGATGACAGTGAAGCC 689  
 DB 69 GAAAGAGAGCTTGAAGAAAGCTTCTCGGCTGGTATGATGACAGTGAAGCC 128  
 QY 690 ACCATAGACAGGAGAGCTATGTTGCGAGTGATGACTTACATGAGTGGATCTCTA 749  
 DB 129 TCCGTGGACCGGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGAT 188  
 QY 750 TGGTGAATGTTTACGAAGAAGAGCTCCCTGAGAGCTGAAAGAAATTCAGGCTATGGA 809  
 DB 189 TGAACATGCTTCAACAAAGAGCTTCTCCATCTGAGAGAGCTGATTCCTATGGA 248  
 QY 810 TGCCGATGAGCTTTTATAGATCATTCGACAGTATGATTTATCTAGCTTTGAGATGC 869  
 DB 249 GAGACAGGCTTTTACAGATCTGCGCCAGTACAAATCATTTGCGCTTTGAGAAAGC 308  
 QY 870 AGTTGTGATGATCATCATCATGAGAGTTCGAGAGCCAGTGAAGTGGGGGTGCTCC 929  
 DB 309 GGTGTGAGAGCTTATCATCATGAGAGTTCGAGAGCCAGTGAAGTGGGGGTGCTCC 368  
 QY 930 TGTATATTAAGAGATGCCCGAGATCATGAGTGGCTTCCAAATGAAAGTCTATTTCT 989  
 DB 369 TGTATATTAAGAGATGCCCGAGATCATGAGTGGCTTCCAAATGAAAGTCTATTTCT 428  
 QY 990 TGTATATTAAGAGATGCCCGAGATCATGAGTGGCTTCCAAATGAAAGTCTATTTCT 1049  
 DB 429 AGTGCAGCCGAGAGAGCTCCCGAGATCATGAGTGGCTTCCAAATGAAAGTCTATTTCT 488  
 QY 1050 TGACAGATTTATGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGAT 1109  
 DB 489 CGACGAGAGTATGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGAT 548  
 QY 1110 ACTTCTGACAGCTCTCAGGAGAAAGAGATGAGAGTCAAGAGCTCAACAGAGACATTA 1169  
 DB 549 GTTGTAAAGAGAGCTGAGAGAGCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGAC 608  
 QY 1170 CATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1229  
 DB 609 CATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 668  
 QY 1230 AAAGAGCTTCAACCCCAAGATGAGAGGAGAGAGATGATGATGATGATGATGATGAT 1289  
 DB 669 AAAGAGCTTCAACCCCAAGATGAGAGGAGAGAGATGATGATGATGATGATGATGAT 728  
 QY 1290 CACAGTGTGCTTTCTCACACTCCGAGCTCAGCTTTGAGCTTTGCGAGAGATGTG 1349  
 DB 729 GAGGATTTGAGCTTTCTCAGGCC--CTGAAAGCGGAGATTTTTCGAGAGATTTG 785  
 QY 1350 GATTTCAGCTTGAACCAATCCAGAAAGAGAGCCAGGACT 1391  
 DB 786 GACGGCCAGTTAGAGAGAGTCCAGAAAGAGGCTAGGGATT 827

Search completed: September 13, 2004, 02:18:00  
 Job time : 2686.88 secs

Db 301 ACCTTCCTCAGCATCTCAGAAATCCATCTGCCATGATGATGGAACTTCTATAAAATAC 360

1./hncse2.Organ.ovary.Vector.pbluescript II KS(+). Site1.  
EcoRI. Site2. NotI. This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was blunt-ended, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pbluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

RESULT 12  
LOCUS BU261379 849 bp mRNA linear EST 26-NOV-2002  
DEFINITION 603374248P1 CSEQCCHNS1 Gallus gallus cDNA clone CHEST286012 5', mRNA  
ACCESSION BU261379  
VERSION BU261379.1 GI:25528988  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 849)  
Boardman,P.E., Sanz-Egualtero,J., Overton,I.M., Burt,D.W., Boscch,E.,  
Fong,W.T., Rickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.,  
A Comprehensive Collection of Chicken cDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)  
JOURNAL 22335534  
MEDLINE 12445392  
PUBMED  
COMMENT Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
FEATURES  
Source Location/Qualifiers  
1..849  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, H1sex"  
/db\_xref="taxon:9031"  
/clone="CHEST286012"  
/dev\_stage="22"  
/lab\_host="DH10B"  
/clone\_id="CSEQCCHNS1"  
/note="Organ: limbs; Vector: pBluescript II KS(+); Site\_1:  
EcoRI; Site\_2: NotI. This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was bluntend, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

Db 183 CATCTGAGAAATTCATCTCCATGATGATGGAACTTTATTAATATCTGCACAGTAC 242  
Qy 844 AAGTTATCTTACCTTTTGAAGATGAGCTTTGATGATGATGATGATGATGATGATG 903  
Db 243 AAGTTATCTTCTCTTTTGAAGATGATGATGATGATGATGATGATGATGATGATG 302  
Qy 904 AGGCCACTGAACTGGGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 963  
Db 303 CGGCTCTGATGCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 362  
Qy 964 CTTCAGATGACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1023  
Db 363 CTTCAGATGACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 422  
Qy 1024 AGTTACATGACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1083  
Db 423 CGCTATGACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 482  
Qy 1084 CTGAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1143  
Db 483 CTGAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 542  
Qy 1144 GTGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 602  
Db 543 GTGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 602  
Qy 1204 GTGCGGCTATATGATGATGATGATGATGATGATGATGATGATGATGATG 1263  
Db 603 GTGCGGCTATATGATGATGATGATGATGATGATGATGATGATGATGATG 662  
Qy 1264 GATGACCACTGATGATGATGATGATGATGATGATGATGATGATGATG 1323  
Db 663 GTTACCATGATGATGATGATGATGATGATGATGATGATGATGATG 719  
Qy 1324 CTTTGAGCTCTTTGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 1383  
Db 720 GCGCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 779  
Qy 1384 CAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1443  
Db 780 TGGGCGGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 838  
RESULT 13  
LOCUS BM090273 523 bp mRNA linear EST 19-NOV-2001  
DEFINITION 505452 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BM090273  
VERSION BM090273.1 GI:17000901  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 523)  
Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Greas,E., Wray,J.B., White,J., Cho,J., Fehsenkrug,S.C.,  
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,  
Chitko-Mckown,C.G., Perlea,G., Holt,I., Karaymcheva,S., Liang,F.,  
Quackenbush,J. and Keeler,J.W.  
Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)  
JOURNAL 21180013  
MEDLINE 11282978  
PUBMED  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 763 4366  
Fax: 402 763 4390  
Email: smith@email.marc.usda.gov

ORIGIN Institute of Mental Health (NIMH)."

Query Match 30.0%; Score 438.4; DB 14; Length 751;  
Best Local Similarity 80.6%; Pred. No. 1.8e-122;  
Matches 527; Conservative 0; Mismatches 121; Indels 6; Gaps 1;

QY 814 GATGGCTTTATAGCAATGACAGTATAGTTATCTGAGCTTTGAGAAATGCACTT 873  
DB 1 GATGCTTTCTACAGACTATGCGCAATTAAGTTATCTGAGCTTTGAGAAATGCGTTC 60  
QY 874 TGTGATGACTACATCACTGAGAAATTTGAGAGCCACTGAACTGAGGAGTATCCCTGTA 933  
DB 61 TGTGAGCATTAATCAACAGAGAAATTTGAGAGCCACTGAACTGAGGAGTATCCCTGTA 120  
QY 934 TATTACGAGATCCCCAGCATCAAGATGAGCTTTCCAGTAACAAAGTCTATTCTTTGTA 993  
DB 121 TATTACGAGATCCCCAGCATCAAGATGAGCTTTCCAGTAACAAAGTCTATTCTTTGTA 180  
QY 994 TCAGAAATTTCTACAGCCAGGAGAACTGCGAAATTAATCAGACAGCTGATTTCTGATGAC 1053  
DB 181 TCGGAATTTCTACAGCCAGGAGAACTGCGAAATTAATCAGACAGCTGATTTCTGATGAC 240  
QY 1054 AGATTGTATGAGGCTTATGATGATGAAAGCTGAAAGGCTGAGATCTTACACAGGACTT 1113  
DB 241 GGGTTGTATGAGGCTTATGATGATGAAAGCTGAAAGGCTGAGATCTTACACAGGACTT 300  
QY 1114 CTACAGGCTCTCAGGAGAGGAAATGGGAGAGCAAGCTCAACAGGAGCAATTAATC 1173  
DB 301 CTACAGGCTCTCAGGAGAGGAAATGGGAGAGCAAGCTCAACAGGAGCAATTAATC 360  
QY 1174 GATGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1233  
DB 361 GATGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
QY 1234 GGGCTTACACCCCAAGATGAGAGGAGAGAGATCCCATCTGATGCTTCCAGAGCCCA 1233  
DB 421 GGGCTTACACCCCAAGATGAGAGGAGAGAGATCCCATCTGATGCTTCCAGAGCCCA 480  
QY 1294 GGTGTTGCTTTCTACACAC-----TCCGAGCTTCAACCTTTGAGCTTTTCCAGAGATG 1347  
DB 481 TGTGTCACCTTCTCATCTCCAGCTTCTCCGAGCTTCAACCTTTGAGCTTTTCCAGAGATG 540  
QY 1348 TGAATTTCCAGCTTTTGAACATCCAAAGAAAGAGCCCAAGCTTAAGTGTGCTGATGAT 1407  
DB 541 TGAATTTCCAGCTTTTGAACATCCAAAGAAAGAGCCCAAGCTTAAGTGTGCTGATGAT 600  
QY 1408 AGGATCAAAAGCTTTTCAATCTCAAGGCTTTGGGGCTTATGATTAAGAGACTGA 1461  
DB 601 AGGATCAAAAGCTTTTCAATCTCAAGGCTTTGGGGCTTATGATTAAGAGACTGA 654

RESULT 11  
BE751825 562 bp mRNA linear EST 25-APR-2001  
LOCUS BE751825  
DEFINITION 204047 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BE751825  
VERSION BE751825.1 GI:10165817  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 562)  
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Casas,E., Wray,J.B., White,J., Cho,J., Fahrenkrug,S.C.,  
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,  
Chitko-Mckown,C.G., Perera,G., Holt,I., Karamyheva,S., Liang,F.,  
Quackenbush,J. and Keefe,J.W.  
Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)

MEDLINE 21180013  
PUBMED 11282978  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAAACAGTATGACCAT  
BACKWARD: GTTCCAGTCAGCAGC  
Plate: 44 row: G column: 22  
Seq primer: ATTTAGTGACCTATG.  
Location/Qualifiers

FEATURES  
source  
1..562  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="MARC 2BOV"  
/note="Vector: pCMV SPORT6; Site\_1: Not; Site\_2: SalI;  
library made from pooled tissue from testis, thymus,  
semiteendonosus muscle, longissimus muscle, pancreas,  
adrenal, and endometrium."

## ORIGIN

Query Match 28.9%; Score 422.6; DB 10; Length 562;  
Best Local Similarity 85.5%; Pred. No. 1.1e-117;  
Matches 483; Conservative 0; Mismatches 79; Indels 3; Gaps 1;

QY 786 GCGTAAATTCAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 845  
DB 1 GTTGAACACCCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 60  
QY 846 GTTATCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 905  
DB 61 GTTATCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120  
QY 906 GGCACCTGAAAGCTGGGGGATGCTCTGATATTAAGAGATCCCGCAGCATCAAGCTGGCT 965  
DB 121 ACCTCTGAAAGCTGGGGGATGCTCTGATATTAAGAGATCCCGCAGCATCAAGCTGGCT 180  
QY 966 TCCAGTAAACAAAGTGTATTTCTTGTATGAGAAATTTCTACCCAGGAACTGGCAAG 1025  
DB 181 TCCAGTAAACAAAGTGTATTTCTTGTATGAGAAATTTCTACCCAGGAACTGGCAAG 240  
QY 1026 TTACATCAGACGATGATTTCTGATGACAGATTTGATGAGGCTTATGATGAGAGCT 1085  
DB 241 CTACATCAGACGATGATTTCTGATGACAGATTTGATGAGGCTTATGATGAGAGCT 300  
QY 1086 GAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1145  
DB 301 GAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360  
QY 1146 GCAAGAGCTCAACAGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1205  
DB 361 GCAAGAGCTCAACAGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
QY 1206 GTGGCTTAATACAGGCTTCCAGAAAGGCTTCCAGAAAGGCTTCCAGAAAGGCTTCCAGAA 1265  
DB 421 GTGGCTTAATACAGGCTTCCAGAAAGGCTTCCAGAAAGGCTTCCAGAAAGGCTTCCAGAA 480  
QY 1266 TACCCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1325  
DB 481 TACCCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 537  
QY 1326 TTTGAGCTTTTGGCAGAGATGAG 1350  
DB 538 TGGGCGCTTTTGGCAGAGATGAG 562

DEFINITION 60180443BFL NCI\_CGAP\_Mams Mus musculus cDNA clone IMAGE:4035584 5',  
 mRNA sequence.  
 ACCESSION BF181631  
 VERSION BF181631.1 GI:11059773  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 945)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Plate: L1AM309 row: k column: 09  
 High quality sequence, stop: 645.  
 Location/Qualifiers  
 1..945  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4035584"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="7 months"  
 /lab\_host="DH10B"  
 /clone\_1lb="NCI CGAP Mams"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI. Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lohar Hennighausen/Robin Humphreys, NIH"

ORIGIN  
 Query March 30.4% Score 444; DB 10; Length 945;  
 Best Local Similarity 81.9%; Pred. No. 3.9e-124; Indels 4; Gaps 4;  
 Matches 560; Conservative 0; Mismatches 120;  
 537 GTTCAGCAGGCGATCCGACTTGCCTTACGAGTCAACCACTTGGAGAGCATGAACTCT 596  
 1 GTTCAGCAGGCGATCCGACTTGCCTTACGAGTCAACCACTTGGAGAGCATGAACTCT 60  
 537 GAAGTCACTCCGATACCTAGTCTTTCGAGTCCAAAACAACTAGAAAAGACTTGC 656  
 61 GAAGTCACTCCGATACCTAGTCTTTCGAGTCCAAAACAACTAGAAAAGACTTGC 120  
 657 TCCGCTGTGTATACAGTCACTGTGACCAACATCAGACAGGAGCATATGTTGC 716  
 121 -CCATGTATATTTTCACTGAGTCCGATCCCAATCAGACAGGAGCATATGTTGC 179  
 717 CGAGTGTGACTTACATCGAGTGCATTCCTATGTTGAATGTTTACGAAACAAAGACT 776  
 180 GGAGGTGATGCGCAATGGAATGATTCCTATGCGAGTGTTCACAAAAGAGATCT 239  
 777 CCTCAGAGCTGAAAATCAGGCTTATGATGCGGATGGCTTTATAGATCATTCG 836  
 240 TCCCTCAGAGCTGAAAATCAGGCTTATGATGCGGATGGCTTTATAGATCATTCG 299  
 837 ACAGTATAGTTTATCCTAGCTTTTGAAATGAGAGTTTGATGACTACATCAGTGA 896  
 300 CCAGTATAGTTTATCCTAGCTTTTGAAATGAGAGTTTGATGACTACATCAGTGA 359  
 897 GTTCGAGGCGCATGAACTGGGGGTAGTCCCGTATATATAGGATCCCGACGATC 956  
 360 GTTCGAGGCGCATGAACTGGGGGTAGTCCCGTATATATAGGATCCCGACGATC 419

957 AGACTGCTTCCAGTAAACAAAGTCTATTCTGTATCAGAAATTTTCTCAACCCAGGGA 1016  
 420 CGACTGCTTCCAGTAAACAAAGTCTATTCTGTATCAGAAATTTTCTCAACCCAGGGA 479  
 1017 ACTGGAGATTACATCAGACGACTGATCTGATGACAGATTGATGAGGCTATGTAGA 1076  
 480 GCTGGGAACTTATATGAGAGCTGATATGACGATGGGCTTGATGAGACTTATGTAGA 539  
 1077 ATGGAAGCTGAAAGGTGATCTTCTTACCGAGCTTCTGACAGTCTTCAGGAAACGGA 1136  
 540 GTGGAAGCTTAAAGGGCAAGATCTTAAACGAGG-CTTCGAGAG-TCTCAATGAACGGGA 597  
 1137 ATGGGAGTCAAGACGTCACACGACATATACATCATCATGATGATGATGATG 1196  
 598 GTGGGGGTGACAGACATTAACAGAGACA-TACATTAAGCTGTTTGAATGATGATG 656  
 1197 CACCAAGTGTGGCTTAATATGAG 1220  
 657 GCCGAGGTGTGGGCAACAGTAG 680

RESULT 10  
 CF748965  
 LOCUS  
 DEFINITION UI-M-HUO-cmj-a-17-0-UI.r1 NIH\_BMAP\_HUO Mus musculus cDNA clone  
 IMAGE:30631600 5', mRNA sequence.  
 ACCESSION CF748965  
 VERSION CF748965.1 GI:37645310  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 751)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Dr. James Lin University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at <http://genome.iowa.edu/distribution/mouse1.html>  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)  
 Seq primer: pyx-5.  
 Location/Qualifiers  
 1..751  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30631600"  
 /tissue\_type="upper head"  
 /dev\_stage="9.5 and 10.5 dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_1lb="NIH\_BMAP\_HUO"  
 /note="Organ: Head; Vector: pyx-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAATGAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National

|    |      |  |       |
|----|------|--|-------|
| Db | 67   | CAGTATTAAGTTTATCTTACGCTTTTGAGAGTCAAGTTGTGATATGACTACATCACTAGAAg | 126   |
| Qy | 898  | TTCTGAGGCGCACTGAACCTGGGGGTAGTCCCTGTATATTACGAGATCCCCAGATCACa    | 957   |
| Db | 127  | TTCTGAGGCGCACTGAACCTGGGGGTAGTCCCTGTATATTACGAGATCCCCAGATCACa    | 186   |
| Qy | 958  | GACTGAGCTTCCAAAGTAAACAAAGTGTCTTCTCTGATCAGATTTTCTCACCACGGGAA    | 101.7 |
| Db | 187  | GACTGAGCTTCCAAAGTAAAGTGTCTTCTCTGATCAGATTTTCTCACCACGGGAA        | 246   |
| Qy | 1018 | CTGGCAAGTTACATCAGACGACTGGATTCTGATGACAGATTGTATGAGGCTTATGAA      | 10777 |
| Db | 247  | CTGGCAAGTTATATCAGACGACTGGATTCTGATGACAGATTGTATGAGGCTTATGAA      | 306   |
| Qy | 1078 | TGSAAGCTAAAGGTGAGATCTCTTACACGAGACTTCTGCACTCTCCAGGAAAGGAAA      | 1137  |
| Db | 307  | TGSAAGCTAAAGGTGAGATCTCTTACACGAGACTTCTGCACTCTCCAGGAAAGGAAA      | 366   |
| Qy | 1138 | TGGGAGTSCAAGACGTCAACCCAGACAAATTACATTCGATCATTTGAGTGTAGTGNGC     | 1197  |
| Db | 367  | TGGGAGTSCAAGACGTCAACCCAGACAAATTACATTCGATCATTTGAGTGTAGTGNGC     | 426   |
| Qy | 1198 | ACCAAGTGTGGGCTTAATATCAAGCTTCAGAAAAGGGCTTACCACCCAAAAGATGGGAG    | 1257  |
| Db | 427  | ACCAAGTGTGGGCTTAATATCAAGCTTCAGAAAAGGGCTTACCACCCAAAAGATGGGAG    | 486   |
| Qy | 1258 | GCAGAAGATACCACCTGAGTGTGCCAGAGGCCACAGTGTGTTGCTTCCACACTCCGG      | 1317  |
| Db | 487  | GCAGAAGATACCACCTGAGTGTGCCAGAGGCCACAGTGTGTTGCTTCCACACTCCGG      | 546   |
| Qy | 1318 | ACTCCACCTTTGAGCTCTTTGCGAGAGATGTGGATTTCCAGCTTTGAACATTCGAGAA     | 13777 |
| Db | 547  | ACTCCACCTTTGAGCTCTTTGCGAGAGATGTGGATTTCCAGCTTTGAACATTCGAGAA     | 606   |
| Qy | 1378 | GAGGCCAAGCACTAAGGTGCGTGGTTGATAGCATCAAAACTTTTCATCTCAGAGTTT      | 14377 |
| Db | 607  | GAGGCCAAGCACTAAGGTGCGTGGTTGATAGCATCAAAACTTTTCATCTCAGAGTTT      | 666   |
| Qy | 1438 | TGGGGCTTGTATTTCAAGACTGA 1461                                   |       |
| Db | 667  | TGGGGCTTGTATTTCAAGACTGA 690                                    |       |

|            |   |            |              |                              |  |  |  |
|------------|---|------------|--------------|------------------------------|--|--|--|
| RESULT 8   | 687 bp  | MRNA       | linear       | EST 27-FEB-2001              |  |  |  |
| LOCUS      | 6024342496671   | NIH_MGC_18 | Homo sapiens | CDNA clone IMAGE:4549985 5', |  |  |  |
| DEFINITION | mRNA sequence.  |            |              |                              |  |  |  |
| ACCESSION  | BCG332068   |            |              |                              |  |  |  |
| VERSION    | BCG332068.1   |            |              |                              |  |  |  |
| KEYWORDS   | EST.  |            |              |                              |  |  |  |
| SOURCE     | Homo sapiens (human)  |            |              |                              |  |  |  |
| ORGANISM   | Homo sapiens  |            |              |                              |  |  |  |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;<br>Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  |            |              |                              |  |  |  |
| AUTHORS    | 1 (bases 1 to 687)  |            |              |                              |  |  |  |
| TITLE      | NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .   |            |              |                              |  |  |  |
| JOURNAL    | National Institutes of Health, Mammalian Gene Collection (MGC)  |            |              |                              |  |  |  |
| COMMENT    | Unpublished (1999)<br>Contact: Robert Strausberg, Ph.D.<br>Email: c9abbs-remail.nih.gov<br>Tissue Procurement: DCTD/DNP/Gazdar<br>CDNA Library Preparation: Ling Hong/Rubin Laboratory<br>CDNA Library Arrayed By: The I.M.A.G.E. Consortium (LIML)<br>DNA Sequencing by: Incyte Genomics, Inc.<br>Clone distribution: MGC clone distribution information can be<br>found through the I.M.A.G.E. Consortium/LIML at:<br><a href="http://image.llnl.gov">http://image.llnl.gov</a><br>plate: L1CM1242 row: d column: 18<br>High quality sequence stop: 635.<br>Location/Qualifiers<br>1..687 |            |              |                              |  |  |  |
| FEATURES   | source  |            |              |                              |  |  |  |

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4549385"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NTH_MGC_18"
/notes="Organ: lung; Vector: pOTS7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAGC(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NTH_MGC Library."

```

| Query Match:           | 38.4%           | Score 561;   | DB 12;     | Length 687; |
|------------------------|-----------------|--|------------|-------------|
| Best Local Similarity: | 95.5%           | Pred. No. 4.3e-160;  |            |             |
| Matches 654;           | Conservative 0; | Mismatches 20;   | Indels 11; | Gaps 7;     |
| QY                     | 519             | GTTCACACTACACTGCGACGTTTGACAGAGCATTCGCCACTTGCCACTACTACCCAACTT | 578        |             |
| Db                     | 2               | GTTTCAACTACACTGCGACGTTTGACAGAGCATTCGCCACTTGCCACTACTACCCAACTT | 61         |             |
| QY                     | 579             | GGAGGACATTGAAGTCCGGAAGTCACTCCGATACCTAGTTCTCTTGGAGTCCAAAACAA  | 638        |             |
| Db                     | 62              | GGAGGACATTGAAGTCCGGAAGTCACTCCGATACCTAGTTCTCTTGGAGTCCAAAACAA  | 121        |             |
| QY                     | 639             | GCTTGAAGAAAAGACTTGGCTCCGCTGGTGTATGTACAGTCAGACTGAGCCACCATCAGA | 698        |             |
| Db                     | 122             | GCTTGAAGAAAAGACTTGGCTCCGCTGGTGTATGTACAGTCAGACTGAGCCACCATCAGA | 181        |             |
| QY                     | 699             | CAGGACACGCTATGTTGGCGAGCTGAGATATTCATCGAGTCGATTCCTATGGGTGAATG  | 758        |             |
| Db                     | 182             | CAGGACACGCTATGTTGGCGAGCTGAGATATTCATCGAGTCGATTCCTATGGGTGAATG  | 241        |             |
| QY                     | 759             | TTTACGAAACAAAGACCTCCCTCAGCAGCTGAAATAATCCAGCCTCTATGATGCGCATG  | 818        |             |
| Db                     | 242             | TTTACGAAACAAAGACCTCCCTCAGCAGCTGAAATAATCCAGCCTCTATGATGCGCATG  | 301        |             |
| QY                     | 819             | CTTTATATGATCATTTGCAAGTATTAAGTTATCTTACGCTTTTGAAATGCAATTTGTGA  | 878        |             |
| Db                     | 302             | CTTTATATGATCATTTGCAAGTATTAAGTTATCTTACGCTTTTGAAATGCAATTTGTGA  | 360        |             |
| QY                     | 879             | TGACTACATCACTGAGAAAGTCTGAGAGCCACTGAAACTGGGGAGTCCCTGTATATTA   | 938        |             |
| Db                     | 361             | TGACTACATCACTGAGAAAGTCTGAGAGCCACTGAAACTGGGGAGTCCCTGTATATTA   | 420        |             |
| QY                     | 939             | CGGATCCCCAGCATCAACAGCTGGCTTCCAGTAACAAATGCTATTTCTTGATCAGA     | 998        |             |
| Db                     | 421             | CGGATCCCCAGCATCAACAGCTGGCTTCCAGTAACAAATGCTATTTCTTGATCAGA     | 480        |             |
| QY                     | 999             | ATTTTCTACCCCGAGGGAATCGCAAGTTTCATACAGACATCGAATTCGATGACAGATT   | 1058       |             |
| Db                     | 481             | ATTTTCTACCCCGAGGGAATCGCAAGTTTCATACAGACATCGAATTCGATGACAGATT   | 539        |             |
| QY                     | 1059            | GTAATGAGGCTATGTAGATGGAAGCTGA-AGGGTGAATCTCTAACAGAGCA-CTTCT    | 1115       |             |
| Db                     | 540             | GTAATGAGGCTATGTAGATGGAAGCTGA-AGGGTGAATCTCTAACAGAGCACTCTCTG   | 599        |             |
| QY                     | 1116            | GACACTCTCAGGG--AACGGAATGGGGAGTGCAGAGAG--TCAACAGAGACAA--TTA   | 1169       |             |
| Db                     | 600             | AACACTCTCAGGGACCGAACHCGAGTGGGAGTGCAGAGAGCTCAACACAGACAAATTAC  | 659        |             |
| QY                     | 1170            | CATCGATCATTTGATGATGTATGTG 1194                               |            |             |
| Db                     | 660             | ATCGGATCATTCGAGTGTATGTG 684                                  |            |             |

RESULT 9  
BF181631

BF181631 945 bp mRNA linear EST 31-OCT-2000

375 Newton Road, 4156 WEBER, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 Reverse.  
 Location/Qualifiers

1..712  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-EJ0-ahk-e-07-0-UI"  
 /tissue\_type="fetal eyes, lens, eye anterior segment,  
 optic nerve, retina, Retina Foveal and Macular, RPE and  
 Choroid"  
 /dev\_stage="fetal and adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_1lb="UI-E-EJ0"  
 /note="Organ: eye; Vector: pT73-Pac (pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-E-EJ0 is a subcloned cDNA library constructed  
 according to Bonaldo, Lennon and Soares, Genome Research,  
 6:791-806, 1996. First strand cDNA synthesis was primed  
 with an oligo-dT primer containing a Not I site. Double  
 stranded cDNA was ligated to an EcoR I adaptor, digested  
 with Not I, and cloned directionally into pT73-Pac  
 vector. The oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (dT)<sub>18</sub> tail. The  
 sequence tags for this library are: fetal eyes,  
 AGAATCAAG; lens, CGATTAGGA; eye anterior segment,  
 AATGCCGAT; optic nerve, CCATTAGGT; retina, CCGCG; Retina  
 Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This  
 library was created for the program, Gene Discovery in the  
 Visual System, supported by National Eye Institute (NEI)."

ORIGIN  
 Query Match 47.5%; Score 694.4; DB 12; Length 712;  
 Best Local Similarity 99.6%; Pred. No. 5.5e-201;  
 Matches 706; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
 QY 505 CCAGGATTAACCTGTTCACTACAGCGACGTTGACGAGGATCCCACTTGCCACTA 564  
 DB 1 CAGGTGATCAGCTGTTCACTACAGCGACGTTGACGAGGATCCCACTTGCCACTA 60  
 QY 565 ACTACCCATACTTGAAGAGCATTGAAGTCTGAAGTCACTCCGATACCTAGTTCCTTTG 624  
 DB 61 ACTACCCATACTTGAAGAGCATTGAAGTCTGAAGTCACTCCGATACCTAGTTCCTTTG 120  
 QY 625 CAGTCCAAAAAACAAGCTTGAAGAAAGCTGCTCCGCTGTATATACAGTCAGACTCT 684  
 DB 121 CAGTCCAAAAAACAAGCTTGAAGAAAGCTGCTCCGCTGTATATACAGTCAGACTCT 180  
 QY 685 GACCCACCATCAGACAGGACAGCTATGTTGAGAGCTGATGACTTACATCGAGTGCAT 744  
 DB 181 GACCCACCATCAGACAGGACAGCTATGTTGAGAGCTGATGACTTACATCGAGTGCAT 240  
 QY 745 TCCATGCTGATGTTTACGAAACAAAGACCTCCCTAGAGCTGAAAATCCAGCTCT 804  
 DB 241 TCCATGCTGATGTTTACGAAACAAAGACCTCCCTAGAGCTGAAAATCCAGCTCT 300  
 QY 805 ATGATGCCAGATGCTTTATAGATCATTCAGACAGTAAAGTTTATCCAGCTTTGAG 864  
 DB 301 ATGATGCCAGATGCTTTATAGATCATTCAGACAGTAAAGTTTATCCAGCTTTGAG 360  
 QY 865 AATGCAATTGTGATGACTACATCACTGAGAAAGTTTGGAGGCCACTGAAACTGGGGGTA 924  
 DB 361 AATGCAATTGTGATGACTACATCACTGAGAAAGTTTGGAGGCCACTGAAACTGGGGGTA 420

QY 925 GTCCCTGTATATTACGATCCCCAGACATCACAGACTGCTTCCAGTAAACAAAAGTCT 984  
 DB 421 GTCCCTGTATATTACGATCCCCAGACATCACAGACTGCTTCCAGTAAACAAAAGTCT 480  
 QY 985 ATCTCTGTATACAAATTTTCTCAACCCAGGGAACGTGGAAGTTATATAGACGATGGAT 1044  
 DB 481 ATCTCTGTATACAAATTTTCTCAACCCAGGGAACGTGGAAGTTATATAGACGATGGAT 540  
 QY 1045 TCTGATGACAGATTGTATAGAGCTTATGATGAAAGCTGMAAGGTGAGATCTTAAC 1104  
 DB 541 TCTGATGACAGATTGTATAGAGCTTATGATGAAAGCTGMAAGGTGAGATCTTAAC 600  
 QY 1105 CACGCACTTCTGACACTCTTCAGGGAACGGAATAGGGAGTCCAGACGACAGGAC 1164  
 DB 601 CACGCACTTCTGACACTCTTCAGGGAACGGAATAGGGAGTCCAGACGACAGGAC 660  
 QY 1165 AATTATATCATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1213  
 DB 661 AATTATATCATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 708

RESULT 7  
 BX348736  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 960)  
 L1,W.B., Gruber,C., Jesse,J. and Polayes,D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 CONTACT: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 4610.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0BA009ZG08.CS00831.1&cluster=4610.r.  
 Contact : Peng Liang Email : fliang@life-tech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0BA009ZG08.CS00831\_1.

# FEATURES

source  
 1..960  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0BD001Y24"  
 /tissue\_type="NEUROBLASTOMA COT 50-NORMALIZED"  
 /clone\_1lb="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT)  
 primer. Five prime and enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 46.3%; Score 676; DB 13; Length 960;  
 Best Local Similarity 99.3%; Pred. No. 2.8e-195;  
 Matches 679; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 778 CCTCAGAGCTGAAATATCCAGCTTATGATGATGCGATGCTTTTATAGATCAATTGCA 837  
 DB 7 CCTCAGAGCTGAAATATCCAGCTTATGATGATGCGATGCTTTTATAGATCAATTGCA 66  
 QY 838 CAGTATAGTTTATCTAGCTTTTGAAGATGCAATTGTGATGACTACATCACTGAGAAG 897

Db 920 GATCCCCACCATGCGGACTGCTTCCAGTAAACGAGCCCACTTCGTGATGGAT 979

Qy 1001 TTCTCACCAGGGAAGTGGCAAGTTAATCATCGAGCACTGGAATTCGTATGACAGATGT 1060

Db 980 TCTCTACCCCAAGAGAGCTGGCCAACTTCATTNGAGACTGATTTATGACATGGATTGT 1039

Qy 1061 ATGAGCCCTATGTAGAAATGGAAGCTGAAGGGTGAATCTCTAACCAAGCACTTCGACAG 1120

Db 1040 ATGAGACCTATGTAGAGTGGAGAACTAAAGGGCAAGATCTCTAACCAAGCGGCTTCGACAG 1099

Qy 1121 CTCCTCAGGAGCAAGGAATGGGGAGTGCAGAGCTCAACAGAGCAAAATTAATCAATGAT 1180

Db 1100 CTCCTCAATGAAACGAGAGTGGGGTGTGCAGAGCACTTAACAGAGCAACTACATGCTGT 1159

Qy 1181 TTGAGTGTATGTGTGCACCAAGGTGGGCTAATATCAGGCTTCAGGAAAG 1233

Db 1160 TTGAGTGTATGTGTGTGCGGAGGGTGTGGGCAAAAGTAGGCTGCAGAGGAGAG 1212

RESULT 5  
AL528778 1201 bp mRNA linear EST 23-MAY-2003  
LOCUS AL528778 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens  
DEFINITION cDNA clone CS0DD001YG24 5-PRIME, mRNA sequence.  
ACCESSION AL528778  
VERSION AL528778.2 GI:31066628  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1201)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On Feb 13, 2001 this sequence version replaced gi:12792271.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr; Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4610.r For more information about this cluster, see  
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DD001BD12QPLcluster=4610.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DD001BD12QPL.  
Location/Qualifiers  
1. 1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DD001YG24"  
/tissue\_type="NEUROBLASTOMA COT 50-NORMALIZED"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dt) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 54.3%; Score 794; DB 9; Length 1201;  
Best Local Similarity 96.8%; Pred. No. 2, 1e-231;  
Matches 813; Conservative 7; Mismatches 18; Indels 2; Gaps 1;

Qy 97 AAGAGGTCATGTTAGCTGGGGAAGTTGAAAGAGAGATTAAAGTTCCAGTTTG 156

Db 307 AACGAGTCATGTTAGCTGGGGAAGTTGAAAGAGAGATTAAAGTTCCAGTTTG 366

Qy 157 CAAGATGACATCAAAAATGAGAGAACCTACGATCTTAATTCATTCTTAAGAAA 216

Db 367 CAAGATGACATCAAAAATGAGAGAACCTACGATCTTAATTCATTCTTAAGAAA 426

Qy 217 GAAGATTGACCTTCAACAGAGAAAAGAAATGGAAATTTGACAGTACCCCATATGCTC 276

Db 427 GAAGATTGACCTTCAACAGAGAAAAGAAATGGAAATTTGACAGTACCCCATATGCTC 486

Qy 277 TGGTGGTCCCCCTGACGGGGAGACTGGAGAGTTAGGCCAATGTGAGCAGATGCTTGT 336

Db 487 TGGTGGTCCCCCTGACGGGGAGACTGGAGAGTTAGGCCAATGTGAGCAGATGCTTGT 546

Qy 337 TTCTTACCATCAACCGGACCTACCTCCATCATGACATGACCAAGCACTTCCTCTAT 396

Db 547 TTCTTACCATCAACCGGACCTACCTCCATCATGACATGACCAAGCACTTCCTCTAT 606

Qy 397 GGTAAGCTTTAATAGATAGTAACTTACCTCTGCTGGAAAGCCCATATGACTGGGCT 456

Db 607 GGTAAGCTTTAATAGATAGTAACTTACCTCTGCTGGAAAGCCCATATGACTGGGCT 666

Qy 457 GTTTTCAATGAGAGTCCCGGAAAACAATTATAGCTCTTCAATAACAGATGATACC 516

Db 667 GTTTTCAATGAGAGTCCCGGAAAACAATTATAGCTCTTCAATAACAGATGATACC 726

Qy 517 TTGTTCACTACACCTGACAGCTTCAGCAGAGGATCCCACTTCCACTACTACCAATAC 576

Db 727 TTGTTCACTACACCTGACAGCTTCAGCAGAGGATCCCACTTCCACTACTACCAATAC 786

Qy 577 TTGAGAGATTTGAAGTCTCTGAAGTCACTCCGATACCTAGTCTCTTTCAGAGTCCAAAAC 636

Db 787 TTGAGAGATTTGAAGTCTCTGAAGTCACTCCGATACCTAGTCTCTTTCAGAGTCCAAAAC 846

Qy 637 AAGCTTGAAGAAAGACTGCTCCGCTGGTATATGATGATGAGTGAAGTGAAGCCACATCA 696

Db 847 AAGCTTGAAGAAAGACTGCTCCGCTGGTATATGATGATGAGTGAAGTGAAGCCACATCA 906

Qy 697 GACAGGAGACGTATGTTGCGAGCTGATGATGATGATGATGATGATGATGATGATGAT 756

Db 907 GACAGGAGACGTATGTTGCGAGCTGATGATGATGATGATGATGATGATGATGATGAT 966

Qy 757 TGTTTACGAAAGAAAGACTGCTCCGCTGGTATATGATGATGATGATGATGATGATGAT 816

Db 967 TGTTTACGAAAGAAAGACTGCTCCGCTGGTATATGATGATGATGATGATGATGATGAT 1026

Qy 817 GCGTTTATAGATCATTTGACAGATATGATATGATATGATATGATATGATATGATATGAT 876

Db 1027 GCGTTTATAGATCATTTGACAGATATGATATGATATGATATGATATGATATGATATGAT 1086

Qy 877 GATGACTACATACAGAGATTTCTGAGAGCCACTGAAACTGGGGGTATGCCCTATAT 936

Db 1087 GATGACTACATACAGAGATTTCTGAGAGCCACTGAAACTGGGGGTATGCCCTATATAT 1144

RESULT 6  
LOCUS BM717062 712 bp mRNA linear EST 28-FEB-2002  
DEFINITION UI-E-E30-ahk-e-07-0-UI.r2 UI-E-E30 Homo sapiens cDNA clone  
UI-E-E30-ahk-e-07-0-UI 5', mRNA sequence.  
ACCESSION BM717062  
VERSION BM717062.1 GI:19030320  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 712)  
AUTHORS Bonaldo, M.F., Lencioni, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (5), 791-806 (1996)  
MEDLINE 97044477  
PMID 8889548  
COMMENT Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa



|            |   |   |      |
|------------|---|---|------|
| Db         | 421   | TGGGAAATGTTTACGAAACAAAGACCTCCCTCAGCAGCTGAAAAATCCAGCCCTATGGA     | 480  |
| Qy         | 810   | TGCGGATGCGTTTATAGATCAATGGACAGATTAAGTTATCTAGCTTTTGAATATGC        | 869  |
| Db         | 481   | TGCGGATGCGTTTATAGATCAATGGACAGATTAAGTTATCTAGCTTTTGAATATGC        | 540  |
| Qy         | 870   | AGTTTGTATACATCACTGAGAGAGTTCTGAGAGCCCACTGAAACTGGGGGTAGTCCC       | 929  |
| Db         | 541   | AGTTTGTATACATCACTGAGAGAGTTCTGAGAGCCCACTGAAACTGGGGGTAGTCCC       | 600  |
| Qy         | 930   | TGTATATTACGATCCGCCAGCATCAAGACTGGCTTCCAAAGTAAACAAAGTCTATTC       | 989  |
| Db         | 601   | TGTATATTACGATCCGCCAGCATCAAGACTGGCTTCCAAAGTAAACAAAGTCTATTC       | 660  |
| Qy         | 990   | TGTATATTACGATCCGCCAGCATCAAGACTGGCTTCCAAAGTAAACAAAGTCTATTC       | 1049 |
| Db         | 661   | TGTATATTACGATCCGCCAGCATCAAGACTGGCTTCCAAAGTAAACAAAGTCTATTC       | 720  |
| Qy         | 1050  | TGACAGATTTGATAGGCGCTATGATGAAATGAAAGCTGAAAGGCTGATCTTAACAGCG      | 1109 |
| Db         | 721   | TGACAGATTTGATAGGCGCTATGATGAAATGAAAGCTGAAAGGCTGATCTTAACAGCG      | 780  |
| Qy         | 1110  | ACTTCTGACAGCTCTCAGGGAAACGGAATGGGAGTGCAGACGTC-AAACAGGACAAAT      | 1168 |
| Db         | 781   | ACTTCTGACAGCTCTCAGGGAAACGGAATGGGAGTGCAGACGTC-AAACAGGACAAAT      | 840  |
| Qy         | 1169  | ACATGAGATGCA-TTGTAGTGTATGTTGTG-CACCAAGTGTGGCTTAATATC-AGGCTTC    | 1225 |
| Db         | 841   | ACATGAGATGCA-TTGTAGTGTATGTTGTG-CACCAAGTGTGGCTTAATATC-AGGCTTC    | 900  |
| Qy         | 1226  | AGGAAAGGCGTTACCAACCAAGATGAGAGGAGGAGATACCACTGAGTTGGCCAG          | 1285 |
| Db         | 901   | AGGAAAGGCGTTACCAACCAAGATGAGAGGAGGAGATACCACTGAGTTGGCCAG          | 960  |
| Qy         | 1286  | AGCCCAAGCTTTTGGC-TTTCTCACCACCTCC-GGACTCCACCTTT--GAGCTTTTGGG     | 1340 |
| Db         | 961   | AGCCCAAGCTTTTGGC-TTTCTCACCACCTCC-GGACTCCACCTTT--GAGCTTTTGGG     | 1020 |
| Qy         | 1341  | AGAGATGATGATTTCCAGCTT   | 1361 |
| Db         | 1021  | AAAAATTTGATTTCCAGCTT  | 1041 |
| RESULT 4   | AY405137  | 1212 bp DNA linear GSS 16-DEC-2003                              |      |
| LOCUS      | AY405137  | Mus musculus Funt10 gene, VIRUTAL TRANSCRIPT, partial sequence, |      |
| DEFINITION | AY405137  | Genomic survey sequence.  |      |
| ACCESSION  | AY405137.1  | GI:39761111   |      |
| VERSION    | GSS.  |   |      |
| KEYWORDS   | Mus musculus (house mouse)  |   |      |
| SOURCE     | Mus musculus  |   |      |
| ORGANISM   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  |   |      |
| REFERENCE  | 1 (bases 1 to 1212)   |   |      |
| AUTHORS    | Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B., Ferrieres,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J., Adams,M.D. and Cargill,M. |   |      |
| TITLE      | Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios  |   |      |
| JOURNAL    | Science 302 (5652), 1960-1963 (2003)  |   |      |
| PubMed     | 14671302  |   |      |
| REFERENCE  | 2 (bases 1 to 1212)   |   |      |
| AUTHORS    | Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B., Ferrieres,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J., Adams,M.D. and Cargill,M. |   |      |
| TITLE      | Direct Submission   |   |      |
| JOURNAL    | Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA   |   |      |

|                       |  |   |                 |                   |
|-----------------------|--|---|-----------------|-------------------|
| COMMENT               | This sequence was made by sequencing genomic exons and ordering them based on alignment. |   |                 |                   |
| FEATURES              | Location/Qualifiers  |   |                 |                   |
| source                | 1..1212  |   |                 |                   |
|                       | /organism="Mus musculus"   |   |                 |                   |
|                       | /mol_type="genomic DNA"  |   |                 |                   |
|                       | /db_xref="taxon:10090"   |   |                 |                   |
|                       | <1..>1212  |   |                 |                   |
| gene                  | /gene="Funt10"   |   |                 |                   |
|                       | /locus_tag="HMC2123"   |   |                 |                   |
| ORIGIN                |  |   |                 |                   |
| Query Match           | 54.8%;   | Score 800.2;  | DB 29;          | Length 1212;      |
| Best Local Similarity | 81.6%;   | Pred. No. 2.7e-233;   |                 |                   |
| Matches               | 925;   | Conservative 0;   | Mismatches 208; | Indels 0; Gaps 0; |
| Qy                    | 101  | AGGTATGTTGTTAGCTGGGGAAGTTGAAAGAGAGTTAAAGTTCCAGTTGCAAG       | 160             |                   |
| Db                    | 80   | AGGTATGTTGTTAGCTGGGGAAGTTGAAAGAGAGTTAAAGTTCCAGTTGCAAG       | 139             |                   |
| Qy                    | 161  | ATGACATCAAAAATGGAGGAAGACCTAGCATCTTAATTCATTTCTTAAGAAAGAG     | 220             |                   |
| Db                    | 140  | ACGACATTAAGATGTGAGAGGAGAGCTTAACACCTGAGGCAATTTCTGAAAAGAG     | 199             |                   |
| Qy                    | 221  | GATTGACCTTCAACAGAAAATGGAAATGGAATGGAACGTAACCCATTATGCTTGT     | 280             |                   |
| Db                    | 200  | CATGGCTTGGCTGGAGGACCAAGTGGATGTGTAGTACCACTCCATCGCTGTGT       | 259             |                   |
| Qy                    | 281  | GGTCCCGCTGAGGGGAGACCTGGAGGTAGGCCAATGTGACAGATGCTTGTCT        | 340             |                   |
| Db                    | 260  | GGTCCCGCTGAGGGGAGACCTGGAGGTAGGCCAATGTGACAGATGCTTGTCT        | 319             |                   |
| Qy                    | 341  | TCACCATCAACCGGACCTACCTCATCATCATGACCAACCAATCTCTTCTATGTA      | 400             |                   |
| Db                    | 320  | TTACCATCAACAGGACCTTCCAGATCATTCGATGACAGGCGCTTCTATCTACGGTA    | 379             |                   |
| Qy                    | 401  | CTGACTTTAATAGATAGTACCTTCTGCTCGGAAACCCATCATGACTGGGCTGTT      | 460             |                   |
| Db                    | 380  | CTGACTTTAATAGATAGTACCTTCTGCTCGGAAACCCATCATGACTGGGCTGTT      | 439             |                   |
| Qy                    | 461  | TTGATGAAGATCCCGGAAACCAATTAATTAAGTCTTTGATTAACAGATTAAGTCT     | 520             |                   |
| Db                    | 440  | TCATGAAGATCCCGGAAACCAATTAATTAAGTCTTTGATTAACAGATTAAGTCT      | 499             |                   |
| Qy                    | 521  | TCATGAAGATCCCGGAAACCAATTAATTAAGTCTTTGATTAACAGATTAAGTCT      | 580             |                   |
| Db                    | 500  | TCATGAAGATCCCGGAAACCAATTAATTAAGTCTTTGATTAACAGATTAAGTCT      | 559             |                   |
| Qy                    | 581  | AGAGCATGGAAGTCTGAGATCACTCCGATACCTAGTCTTGGAGTCAAAAACAAGC     | 640             |                   |
| Db                    | 560  | AGAGCATGGAAGTCTGAGATCACTCCGATACCTAGTCTTGGAGTCAAAAACAAGC     | 619             |                   |
| Qy                    | 641  | TTAGAAAAGACTTGGCTCGCTGGTGTATGTAAGTCAAGTCAAGTGTGACCCAGATGAGA | 700             |                   |
| Db                    | 620  | TTAGAAAAGACTTGGCTCGCTGGTGTATGTAAGTCAAGTCAAGTGTGACCCAGATGAGA | 679             |                   |
| Qy                    | 701  | GGGACACTATGTTGGGAGCTGATGATTAATGAGTGCATTCCTATGTTGTAATGT      | 760             |                   |
| Db                    | 680  | GGGACACTATGTTGGGAGCTGATGATTAATGAGTGCATTCCTATGTTGTAATGT      | 739             |                   |
| Qy                    | 761  | TACGAAACAAGACTTCCCTCAGCACTGAAAAAATCAAGCTCTATGATGCGATGCT     | 820             |                   |
| Db                    | 740  | TACGAAACAAGACTTCCCTCAGCACTGAAAAAATCAAGCTCTATGATGCGATGCT     | 799             |                   |
| Qy                    | 821  | TTATATGATCATTTGACAGATTAATGATTAATGATTAATGATTAATGATTAATGAT    | 880             |                   |
| Db                    | 800  | TTATATGATCATTTGACAGATTAATGATTAATGATTAATGATTAATGATTAATGAT    | 859             |                   |
| Qy                    | 881  | ACTACATCACTGAGAGTTCTGGAGGCACTGAAACTGAGGAGTACCTGATATTAG      | 940             |                   |
| Db                    | 860  | ACTACATCACTGAGAGTTCTGGAGGCACTGAAACTGAGGAGTACCTGATATTAG      | 919             |                   |
| Qy                    | 941  | GATCCCGGACATCAAGACTGAGCTTCAAGTAAACAAAGTCTATTTGATTAAGAAAT    | 1000            |                   |

Db 140 ATGACATACAAAATGAGAGAACCTACGATCTTATTCATTCTTAAGAAAG 199  
 QY 221 GATTGACCTTCAACAGAGAAAGAAATGGAAATGGAGACGTAACCCCTTATGCTGGT 280  
 Db 200 GATTGACCTTCAACAGAGAAAGAAATGGAAATGGAGACGTAACCCCTTATGCTGGT 259  
 QY 281 GGTCCCGCTGACGGGGAGAGCTGGAGGTTAGGCAATGTGAGACAGATCTTGTCT 340  
 Db 260 GGTCCCGCTGACGGGGAGAGCTGGAGGTTAGGCAATGTGAGACAGATCTTGTCT 319  
 QY 341 TCACCATCAACCGGACCTTACCTCATCATCATATGACCAAGACATTCCTCTTATGTA 400  
 Db 320 TCACCATCAACCGGACCTTACCTCATCATCATATGACCAAGACATTCCTCTTATGTA 379  
 QY 401 CTGACTTAAACATAGATAGTACTTACCTGCTCCGGAAGGCCATGACTGGGCTGTT 460  
 Db 380 CTGACTTAAACATAGATAGTACTTACCTGCTCCGGAAGGCCATGACTGGGCTGTT 435  
 QY 461 TTGATGAAGAGTCCCGGAAAAAATATATATAGCTCTTTTATTAACAGATTAACCTTGT 520  
 Db 440 TTGATGAAGAGTCCCGGAAAAAATATATATAGCTCTTTTATTAACAGATTAACCTTGT 499  
 QY 521 TCACATCACTGCGACAGTTCAGAGGCAATCCCACTTCCCACTAATCCCAATCTGG 580  
 Db 500 TCACATCACTGCGACAGTTCAGAGGCAATCCCACTTCCCACTAATCCCAATCTGG 559  
 QY 581 AGAGCATGTAAGTCTGAAAGTCACTCCGATACCTAGTCTTTGCAATCAAAAACAGC 640  
 Db 560 AGAGCATGTAAGTCTGAAAGTCACTCCGATACCTAGTCTTTGCAATCAAAAACAGC 619  
 QY 641 TTGAAAAAAGACTGCTCCGCTGGTGTATGTACAGTCTGTGACCAACCATCAACA 700  
 Db 620 TTGAAAAAAGACTGCTCCGCTGGTGTATGTACAGTCTGTGACCAACCATCAACA 679  
 QY 701 GGAGCAGTATGTTGCGAGCTGATGATCATCATGAGAGTCCATGCTATGATGATGTT 760  
 Db 680 GGAGCAGTATGTTGCGAGCTGATGATCATCATGAGAGTCCATGCTATGATGATGTT 739  
 QY 761 TACGAAACAAAGCTCCCTCAGCAGCTGAAAAATCCAGCTCTATGATGCCATGCT 820  
 Db 740 TACGAAACAAAGCTCCCTCAGCAGCTGAAAAATCCAGCTCTATGATGCCATGCT 799  
 QY 821 TTTATGAGATCATGACAGTATAGTTATCTCTGCTTTGAGATGCAATGATGATG 880  
 Db 800 TTTATGAGATCATGACAGTATAGTTATCTCTGCTTTGAGATGCAATGATGATG 859  
 QY 881 ACTACATCACTGAGAACTCTGAGAGCCACTGAAAATGAGGAGTCTGCTATATTAAG 940  
 Db 860 ACTACATCACTGAGAACTCTGAGAGCCACTGAAAATGAGGAGTCTGCTATATTAAG 919  
 QY 941 GATCCCCCAGCATCAAGACTGGCTTCCAAAGTAAAGAAAGTCTATCTTGTATCAAGAT 1000  
 Db 920 GATCCCCCAGCATCAAGACTGGCTTCCAAAGTAAAGAAAGTCTATCTTGTATCAAGAT 979  
 QY 1001 TTTCTACCCCAAGGAACTGCGAAGTTACATCAGACGATGATGATGATGATGATG 1060  
 Db 980 TTTCTACCCCAAGGAACTGCGAAGTTACATCAGACGATGATGATGATGATGATG 1039  
 QY 1061 ATGAGGCTTATGTAGATGAAAGCTGAAAGGTTAGATCTTAAACAGGACTTTGACAG 1120  
 Db 1040 ATGAGGCTTATGTAGATGAAAGCTGAAAGGTTAGATCTTAAACAGGACTTTGACAG 1099  
 QY 1121 CTCTCAGGAGCGAAATGGGAGTGAAGAGTCAACAGAGACATTAATCATGATGAT 1180  
 Db 1100 CTCTCAGGAGCGAAATGGGAGTGAAGAGTCAACAGAGACATTAATCATGATGAT 1159  
 QY 1181 TTGAGTGTATGTGTGACCAAGAGTGTGGCTAATATCAAGGCTTCAAGAAAAAG 1233  
 Db 1160 TTGAGTGTATGTGTGACCAAGAGTGTGGCTAATATCAAGGCTTCAAGAAAAAG 1212

RESULT 3  
 BM559087

LOCUS BM559087 1041 bp mRNA linear EST 20-FEB-2002  
 DEFINITION AGENCOURT 6562690 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:5551224  
 5' mRNA Sequence.  
 ACCESSION BM559087  
 VERSION BM559087.1 GI:18802381  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1041)  
 NIH-MGC http://mgi.nci.nih.gov/.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgabbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LIML at:  
 http://image.liml.gov  
 Plate: LIML2264 row: k column: 01  
 High quality sequence stop: 645.  
 Location/Qualifiers  
 1..1041  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5551224"  
 /tissue\_type="retinoblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_id="NIH\_MGC\_67"  
 /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: oligo dt.  
 Average insert size 1.75 kb. Library constructed by Life  
 Technologies."

ORIGIN  
 Query Match 62.3%; Score 910.4; DB 12; Length 1041;  
 Best Local Similarity 96.5%; Pred. No. 4e-267;  
 Matches 1005; Conservative 0; Mismatches 27; Indels 9; Gaps 7;

QY 330 TECTTGTCTTCCACCATCAACCGGACCTTACCTCATCATCATGACCAAGCATTTCT 389  
 Db 1 TECTTGTCTTCCACCATCAACCGGACCTTACCTCATCATCATGACCAAGCATTTCT 60  
 QY 390 CTCTATGATCTGACCTTAAACATAGATAGTACTCTGCTCGGAAAGCCATCATGA 449  
 Db 61 CTCTATGATCTGACCTTAAACATAGATAGTACTCTGCTCGGAAAGCCATCATGA 120  
 QY 450 CTGGCTGTCTTTCATGAGAGTCCCGAAAAACATTAATAGCTCTTTATTAACCAAGT 509  
 Db 121 CTGGCTGTCTTTCATGAGAGTCCCGAAAAACATTAATAGCTCTTTATTAACCAAGT 180  
 QY 510 GATTACCTTGTCAACTACATGCTGCAAGTTCAGAGGACATTCACATGCTGCACTAATAC 569  
 Db 181 GATTACCTTGTCAACTACATGCTGCAAGTTCAGAGGACATTCACATGCTGCACTAATAC 240  
 QY 570 CCATATCTTGAGAGCATGTAAGTCTGAAAGTCACTCGATACCTAGTCTTCTTGAGTGC 629  
 Db 241 CCATATCTTGAGAGCATGTAAGTCTGAAAGTCACTCGATACCTAGTCTTCTTGAGTGC 300  
 QY 630 CAAAAACAAGCTTGAAGAAAGAACTTGTCTCGCTGCTGTATGTAAGTCAAGTGAACCC 689  
 Db 301 CAAAAACAAGCTTGAAGAAAGAACTTGTCTCGCTGCTGTATGTAAGTCAAGTGAACCC 360  
 QY 690 ACCATCAGACAGGAGCAGTATGTTCCGAGCTGATGATGATGATGATGATGATGATG 749  
 Db 361 ACCATCAGACAGGAGCAGTATGTTCCGAGCTGATGATGATGATGATGATGATGATG 420  
 QY 750 TGTGATATGTTTACGAAACAAAGACCTTCCTCAGCAGCTGAAAAATCCAGCTCTATGGA 809

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.

FEATURES Location/Qualifiers

source 1..1212  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
<1..>1212  
/gene="FUT10"  
/locus\_tag="HOM2123"

## ORIGIN

Query Match 77.4%; Score 1131.4; DB 29; Length 1212;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

101 AGTCATGCTGAGCTGGAGAGTTTGAAGAAGGTTTAAAGTTCACAGTTGCAAG 160  
80 AGTCATGCTGAGCTGGAGAGTTTGAAGAAGGTTTAAAGTTCACAGTTGCAAG 139  
161 ATGACATACAAAATGAGAGACCTTACGATCTTAATTCATTTCTTAAGAAAG 220  
140 ATGACATACAAAATGAGAGACCTTACGATCTTAATTCATTTCTTAAGAAAG 199  
221 GATTGACCTTCAAGAGAAAGAAATGGGAATTGGACAGCTACCCCATTAAGCTG 280  
200 GATTGACCTTCAAGAGAAAGAAATGGGAATTGGACAGCTACCCCATTAAGCTG 259  
281 GATCCCGCTGACGCGGAGAGCTGGAGGTTAGGCCAATGAGAGAGATGCTTGTCT 340  
260 GATCCCGCTGACGCGGAGAGCTGGAGGTTAGGCCAATGAGAGAGATGCTTGTCT 319  
341 TCACCATACCCGACCTTACCTCATATCATATGACCAAGATCTCTTCTTATG 400  
320 TCACCATACCCGACCTTACCTCATATCATATGACCAAGATCTCTTCTTATG 379  
401 CTGACTTTAATATGATAGTCTTACCTGCTCGGAAAGCCCATCATGATGGCTGTT 460  
380 CTGACTTTAATATGATAGTCTTACCTGCTCGGAAAGCCCATCATGATGGCTGTT 439  
461 TTCATGAAGAGTCCCGGAAACATTTAAGCTTTTCAATAACAGATACCTTGT 520  
440 TTCATGAAGAGTCCCGGAAACATTTAAGCTTTTCAATAACAGATACCTTGT 499  
521 TCACCTACCTGACGAGCTTACGAGGATCCCACTGGCAGTACCTACCAATATG 580  
500 TCACCTACCTGACGAGCTTACGAGGATCCCACTGGCAGTACCTACCAATATG 559  
581 AGAGCATTGAAGTCTGAGGATCCGATACCTTCTTCAATGATCGCAAAACAG 640  
560 AGAGCATTGAAGTCTGAGGATCCGATACCTTCTTCAATGATCGCAAAACAG 619  
641 TTAAGAAAAGCTTGCCTGCTGTATGATGATGATGATGATGATGATGATGAT 700  
620 TTAAGAAAAGCTTGCCTGCTGTATGATGATGATGATGATGATGATGATGAT 679  
701 GGGACAGCTAGTGGGAGCTGATGATGATGATGATGATGATGATGATGATGAT 760  
680 GGGACAGCTAGTGGGAGCTGATGATGATGATGATGATGATGATGATGATGAT 739  
761 TACGAAACAAAGACCTCTCTGACGAGTGAATAATCCAGCTTATGATGATGAT 820  
740 TACGAAACAAAGACCTCTCTGACGAGTGAATAATCCAGCTTATGATGATGAT 799  
821 TTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 880  
800 TTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 859  
881 ACTACATCATGAGAGGTTCTGAGGCGCATGAACTGGGAGGATGCTCTGATAT 940  
860 ACTACATCATGAGAGGTTCTGAGGCGCATGAACTGGGAGGATGCTCTGATAT 919

941 GATCCCCCAGCATACAGACTGCGCTTCCAAAGTAACAAAAGTGTATTTGATCAGAT 1000  
920 GATCCCCCAGCATACAGACTGCGCTTCCAAAGTAACAAAAGTGTATTTGATCAGAT 979  
1001 TTTCTACCCCGAGGAGATCGGAGATGATGATGATGATGATGATGATGATGAT 1060  
980 TTTCTACCCCGAGGAGATCGGAGATGATGATGATGATGATGATGATGATGAT 1039  
1061 ATGAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1120  
1040 ATGAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1099  
1121 CTCTCAGGAAACCGAAATGGGAGTGCAGAGAGCTCAACAGAGCAATTAATGAT 1180  
1100 CTCTCAGGAAACCGAAATGGGAGTGCAGAGAGCTCAACAGAGCAATTAATGAT 1159  
1181 TTGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1233  
1160 TTGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1212

RESULT 2  
AY405136 1212 bp DNA linear GSS 16-DEC-2003  
LOCUS  
DEFINITION  
genomic survey sequence.  
ACCESSION  
AY405136.1 GI:39761110  
KEYWORDS  
GSS.  
SOURCE  
Pan troglodytes (chimpanzee)  
ORGANISM  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.  
REFERENCE  
1 (bases 1 to 1212)  
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejaival,A.,  
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,  
Fertler,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
TITLE  
JOURNAL  
Science 302 (5652), 1960-1963 (2003)  
PUBMED  
14671302  
REFERENCE  
2 (bases 1 to 1212)  
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejaival,A.,  
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,  
Fertler,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
Direct Submission  
JOURNAL  
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT  
This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
Location/Qualifiers  
source 1..1212  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
<1..>1212  
/gene="FUT10"  
/locus\_tag="HOM2123"

ORIGIN

Query Match 76.5%; Score 1117.4; DB 29; Length 1212;

Best Local Similarity 98.9%; Pred. No. 0;

Matches 1121; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

101 AGTCATGCTGAGCTGGAGAGTTTGAAGAAGGTTTAAAGTTCACAGTTGCAAG 160  
80 AGTCATGCTGAGCTGGAGAGTTTGAAGAAGGTTTAAAGTTCACAGTTGCAAG 139  
161 ATGACATACAAAATGAGAGACCTTACGATCTTAATTCATTTCTTAAGAAAG 220

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: September 12, 2004, 21:42:25 ; Search time 2684.88 Seconds

(without alignments)  
16249.761 Million cell updates/sec

Title: US-10-080-960-3

Perfect score: 1461

Sequence: 1 atgaagtcacggccctcc.....gctatgattcaaggactga 1461

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues 55026578

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: em\_estba:\*  
2: em\_estbun:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hiv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_vhl:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match | Length DB | ID | Description         |
|------------|--------|-------------|-----------|----|---------------------|
| 1          | 1131.4 | 77.4        | 1212      | 29 | AY405135 Homo sapi  |
| 2          | 1117.4 | 76.5        | 1212      | 29 | AY405136 Pan trogl  |
| 3          | 910.4  | 62.3        | 1041      | 12 | BMS559087 AGENCOURT |
| 4          | 800.2  | 54.8        | 1212      | 29 | AY405137 Mus muscu  |

|    |       |      |      |    |          |                    |
|----|-------|------|------|----|----------|--------------------|
| 5  | 794   | 54.3 | 1201 | 9  | AL528778 | AL528778           |
| 6  | 694.4 | 47.5 | 712  | 12 | BW717062 | BW717062 UT-E-BJO- |
| 7  | 676   | 46.3 | 960  | 13 | BX348736 | BX348736 BX348736  |
| 8  | 561   | 38.4 | 687  | 12 | BG332068 | BG332068 602432466 |
| 9  | 444   | 30.4 | 945  | 10 | BF181631 | BF181631 601804438 |
| 10 | 438.4 | 30.0 | 751  | 14 | CF748965 | CF748965 UT-M-HD-  |
| 11 | 422.6 | 28.9 | 562  | 10 | BE751825 | BE751825 204047 MA |
| 12 | 415.2 | 28.4 | 849  | 13 | BH261379 | BH261379 603374248 |
| 13 | 390   | 26.7 | 523  | 12 | BMO90273 | BMO90273 505452 MA |
| 14 | 388.8 | 26.6 | 728  | 13 | BU449301 | BU449301 603764226 |
| 15 | 350.6 | 24.0 | 839  | 13 | BX873167 | BX873167 BX873167  |
| 16 | 349.4 | 23.9 | 522  | 9  | AL599664 | AL599664 DKFZP313C |
| 17 | 347.8 | 23.8 | 555  | 10 | BE665357 | BE665357 154293 MA |
| 18 | 346.2 | 23.7 | 724  | 13 | BH300928 | BH300928 60373330  |
| 19 | 328.4 | 22.5 | 871  | 12 | BG761716 | BG761716 602717910 |
| 20 | 328   | 22.5 | 771  | 14 | CF738017 | CF738017 UT-M-HD-  |
| 21 | 325.4 | 22.3 | 672  | 13 | BH391962 | BH391962 603510787 |
| 22 | 323.4 | 22.1 | 820  | 13 | BU440303 | BU440303 603019157 |
| 23 | 316.2 | 21.6 | 707  | 13 | BU476024 | BU476024 603846742 |
| 24 | 315.2 | 21.6 | 595  | 13 | BU464928 | BU464928 603776201 |
| 25 | 307   | 21.0 | 1031 | 13 | BU449854 | BU449854 603218748 |
| 26 | 297.2 | 20.3 | 696  | 29 | BX150379 | BX150379 Danio rer |
| 27 | 296.4 | 20.3 | 431  | 14 | CB546903 | CB546903 AMGNMUC:N |
| 28 | 293   | 20.1 | 490  | 9  | AT140637 | AT140637 qd37a06.x |
| 29 | 287   | 19.6 | 501  | 28 | AQ384477 | AQ384477 HS 3118.B |
| 30 | 285.2 | 19.5 | 918  | 13 | BQ725661 | BQ725661 AGENCOURT |
| 31 | 282.2 | 19.3 | 938  | 13 | BU514180 | BU514180 AGENCOURT |
| 32 | 270.2 | 18.5 | 806  | 12 | BI739095 | BI739095 603361351 |
| 33 | 269.2 | 18.4 | 985  | 10 | BE745086 | BE745086 601576194 |
| 34 | 267.6 | 18.3 | 912  | 13 | BU128147 | BU128147 603114758 |
| 35 | 264.4 | 18.1 | 791  | 14 | CB988679 | CB988679 AGENCOURT |
| 36 | 250.6 | 17.2 | 865  | 13 | BQ893273 | BQ893273 AGENCOURT |
| 37 | 234.4 | 16.0 | 412  | 12 | BM254238 | BM254238 515603 MA |
| 38 | 230.2 | 15.8 | 292  | 9  | A1693529 | A1693529 wd43f05.x |
| 39 | 221.6 | 15.2 | 239  | 10 | BF334244 | BF334244 RCL-CT024 |
| 40 | 217.6 | 14.9 | 1044 | 13 | BX348997 | BX348997 BX348997  |
| 41 | 216.2 | 14.8 | 619  | 13 | BU610965 | BU610965 UT-M-FC-  |
| 42 | 212.6 | 14.6 | 892  | 13 | BU380453 | BU380453 603861250 |
| 43 | 212.2 | 14.5 | 924  | 12 | BI755527 | BI755527 603027388 |
| 44 | 210.4 | 14.4 | 744  | 13 | BX856607 | BX856607 BX856607  |
| 45 | 207.2 | 14.2 | 798  | 14 | CF738105 | CF738105 UT-M-HD-  |

## ALIGNMENTS

|            |  |             |     |        |                 |
|------------|--|-------------|-----|--------|-----------------|
| RESULT 1   | AY405135   | 1212 bp     | DNA | linear | GSS 16-DEC-2003 |
| LOCUS      | AY405135   |             |     |        |                 |
| DEFINITION | Homo sapiens FUT10 gene, VIRTUAL TRANSCRIPT, partial sequence.   |             |     |        |                 |
| ACCESSION  | AY405135.1   | GI:38761109 |     |        |                 |
| VERSION    | AY405135.1   | GI:38761109 |     |        |                 |
| KEYWORDS   | GSS.   |             |     |        |                 |
| SOURCE     | Homo sapiens (human)   |             |     |        |                 |
| ORGANISM   | Homo sapiens   |             |     |        |                 |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  |             |     |        |                 |
| AUTHORS    | Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Clevello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,D.J., Adams,M.D. and Cargill,M. |             |     |        |                 |
| TITLE      | Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios   |             |     |        |                 |
| JOURNAL    | Science 302 (5652), 1960-1963 (2003)   |             |     |        |                 |
| PUBMED     | 14671302   |             |     |        |                 |
| REFERENCE  | 2 (bases 1 to 1212)  |             |     |        |                 |
| AUTHORS    | Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Clevello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,D.J., Adams,M.D. and Cargill,M. |             |     |        |                 |
| TITLE      | Direct Submission  |             |     |        |                 |

Db 786 TGCCCGGGACCGCCTAATCT---GCGCCGCGCGGTGCTCCGCCATGGAACGCGCGAGT 842  
QY 638 AGCTTAAAAAAGACTTCTCCGCTGCTGTATGTACAGTCAAGCTGTGACCCACCATCAG 697  
Db 843 GGGCCCCCGCGGTACGCGCCGCTCTCTATCTGCAGTCACTGCACTGCGACGCGCG 902  
QY 698 ACAAGGACAGCTATGTTTCGAGCTGATGACTTACATCGAGGTGATTCCTATGTTGAAT 757  
Db 903 ACCGGACCGCTACGTGCGGAGCTCATGGGCACATCCCGGTAGACTCTCTACGGGAAAT 962  
QY 758 GTTACGAAACAAAGACTTCCT-----CAGCAGCTGAAAAATCCAGCCTCTATG 808  
Db 963 GCCGTGAGAAATCGGAGCTGCTCACCAGCGGCTACAGACACAGCCAGGCCACACC 1022  
QY 809 ATGCCAGTGCCTTTTATAGATCATTCACAGTATAAGTTTATCTTACGCTTTTGAATG 868  
Db 1023 AGGATCCAGAGCTTTGGCTTTCTGTCCCGTATTAAGTTCCACTTGCGCTGAAAAATG 1082  
QY 869 CAGTTTGTATGACTATCATCTGAGAAAGTTCTGAGGCCACTGAAACTGGGGGTAGTCC 928  
Db 1083 CCATCTGTAAAGACTATGATGACGAAAAAACTGTGGCGTCCCATGCACTGGGCGCTGTGC 1142  
QY 929 CTGTATATTACGGATCCGCCAGCATCAAGACTGCTTCCAAAGTAAACAAAAGTGTATTTC 988  
Db 1143 CCGTGTACCGGGTTCCTCCCTCTGTGTAGGGAAGTGAATGCCGAACATTCATCCGTGATCC 1202  
QY 989 TTGTATCAGAAATTTTCTACCCCAAGGAACTGCGAAAGTTACATCAGACACTGATTTCTG 1048  
Db 1203 TGATTGATGATTTTGAAGTCTCTCAGAAAGCTGGCAGAGTTTATTTGACTTTCTGACAA 1262  
QY 1049 ATGACAGATTGTATGAGGCCATGTAGATGGAAGCTGAAGGTTAGATCTTAAACCAAGC 1108  
Db 1263 ATGATGAGGAGTATATGAAATACCTGSCATACAAAGCACTGSGGGCATCACCAACCAAT 1322  
QY 1109 GACTTCTGACAGCTCTGAGGAAACGAAATGGGAGTGCAGAAGCTCAACCAAGGACAAAT 1168  
Db 1323 TTCTTCTGTAGATAGTGTAGAGCATGGGAGTGGGAGTGAATGATCCTTTGCTGCTTAAT 1382  
QY 1169 ACATGATGCAATTGAGTGTATGTTGTGACCAAGTGTGGGCT 1212  
Db 1383 AACTCAAGGCTTCGAGTGTTCGTGTGACTACGAATGGGCT 1426

Search completed: September 13, 2004, 05:00:11  
Job time : 528.98 secs

```

; PRIOR APPLICATION NUMBER: US 09/842,528
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: PCT/US01/40607
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/199,500
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/882,836
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19543
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/211,730
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/882,872
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19153
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/212,077
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-184-648-21

Query Match      12.4%; Score 180.8; DB 13; Length 1479;
Best Local Similarity 51.8%; Pred. No. 6.2e-49;
Matches 489; Conservative 0; Mismatches 442; Indels 13; Gaps 3;

QY 278 GGTGTCCTCCCGCTGAGCGGGGAGACTGGAGGTAGGCCAATGTCAGAGATGCTTTGTT 337
Db 252 GCATTCCCGCACTTCCCGGAGACTCGAGCGCA-TCGAGTGTGCGCGCGCGGTGCG 310
QY 338 TCTTACCATCAACCGGACCTTCTCATCATCAATGACCAAGCAATCTCTTTATG 397
Db 311 TGCGCTCCCGAAACCGCCAGCGCTAGGAGACTCGCGGACGCGCGCTGCTTCTACG 370
QY 398 GTACTGACTTAACTAGATAGTAGTCTGCTGCTCGGAAAGCCCATCATAGTGGGCTG 457
Db 371 GCACAGACTCCCGCGGTGGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGC 430
QY 458 TTTTATGAAAGAGTCCCGGAAACATTAATTAAGCTTTTATTAACCAAGTATTACT 517
Db 431 TCTCCACGAGGAGTGGCCCTCAACAACTTCTGTGAGCGACGCGCGCGCGCGCGC 490
QY 518 TGTTCACACTACATGCGACAGTTGACAGGACATTCCTGCTGCTGCTGCTGCTGCT 577
Db 491 TCTTCAATCTTACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 550
QY 578 TGGAGAGATTGAGTCTGAAGTCACTCCGATCTTCTGCTTCTGCTGCTGCTGCTGCT 637
Db 551 TGCCCGGACCGGCTATCT---GCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 607
QY 638 AGCTTGAAGAAAGACTTGTCTCGCTGTGTATGTACAGTCAAGCTGTGACCCACATCAG 697
Db 608 GGGCGCGCGCGCTACGCGCGCGCTCTATCTGCACTGCACTGCGGACATGCGCACGCG 667
QY 698 ACAGGACAGCTATGTTGCGAGCTGATGACTTATCAAGAGTGGATTCCATGAGTAAAT 757
Db 668 ACCGGGACCGCTACGTCGCGGAGCTCAATGCGCCACATCCGCTGAGCTCTTCAAGGAAAT 727
QY 758 GTTTACGAAACAAAGACTCTCT-----CAGCAGCTGAAAAATTCAGCCCTCTATGG 808
Db 728 GCGTGCAGATCGGGAGCTGCTTACCGCGCGCTACAGACACAGCCACGCGCACCG 787
QY 809 ATGCCAGATGCTTTATAGGATCATTTGACAGTAAAGTTATGCTTCTGCTTTTGAAGATG 868
Db 788 AGGATCGAAGCTCTTGGCTTTCTTGTCCCGCTATAGTTCACTTGGCCCTGGAAGATG 847
QY 869 CAGTTTGTAGTACTACATCACTGAGAAATTCTGAGGCGCACTGAAATGAGGAGTATGCC 928
Db 848 CCATCTGTAAAGACTACATGACGAAAAATGTGTGCGCTCCCATGCACTCTGGGCGCTGTGC 907
```

```

QY 929 CTGTATATTACGATCCCGGAGATCAGACATGCTTCCAGTAACAAAGTGTATTTC 988
Db 908 CCGTGTACCGCGGTTCTCCCTCTGTGAGGAGCTGATGCGCAACATATCTCCGATCC 967
QY 989 TTGTATCAGATTTTCTTCAACCCGAGGAACTGGCAAGTTACATCAAGACTGATTTCTG 1048
Db 968 TGATGATATTGATGCTCTCTCAAGAGCTGGCAGATTATGACTTTCTGACAAAGA 1027
QY 1049 ATGACAGATTGATAGGCTTATGAGATGAGAACTGAAGGCTGAAGTCTTCAACGAC 1108
Db 1028 ATGATGAGAGTATGAAATATCTGCGATACAGCAACCTGGGGGATCAACCAACAT 1087
QY 1109 GACTTGTACAGCTCTCAGGAAACGAAATGGGAGTGCACAGACGTCAACGAGCAATT 1168
Db 1088 TTCTTGTGATAGTCTGAAGCATCGGAGTGGGAGTGAATGATCTTGTGCTTAAT 1147
QY 1169 ACATCATGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1212
Db 1148 ACCTCAACGCTTCAGATGTTTGTCTGATGACGAAGTGGCT 1191

RESULT 15
US-09-844-948-1
; Sequence 1, Application US/09844948
; Publication No. US20030119161A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 32132. A NOVEL FUCOSYLTRANSFERASE FAMILY
; FILE REFERENCE: 10448-048001
; CURRENT APPLICATION NUMBER: US/09/844,948
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: US 60/200,604
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2557
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (236)...(1711)
US-09-844-948-1

Query Match      12.4%; Score 180.8; DB 10; Length 2557;
Best Local Similarity 51.8%; Pred. No. 9e-49;
Matches 489; Conservative 0; Mismatches 442; Indels 13; Gaps 3;

QY 278 GGTGTCCTCCCGCTGAGCGGGGAGACTGGAGGTAGGCCAATGTCAGAGATGCTTTGTT 337
Db 487 GCTATTTCCCACTTCCCGGAGACTCGAGCGCA-TGAGTGTGCGCGCGCGGTGCG 545
QY 338 TCTTACCATCAACCGGACCTTCTCATCATCAATGACCAAGCAATCTCTTTATG 397
Db 546 TGGCGTCCCGAAGCGCGGAGCGCTGAGGAGACTGCGGAGCGCGCGCGCTTCTTACG 605
QY 398 GTACTGACTTTAAATAGATAGTACTTCTGCTCGGAAAGCCCATGATGAGTGGCTG 457
Db 606 GCACAGACTTCCGCGCTGCGCGCGCGCGCTGCGCGCGCTGCGGCGACACAGACTGGCGCG 665
QY 458 TTTTCAATGAGAGTCCCGGAAACATTAATAGCTCTTCTTAACCAAGTATTAAGCT 517
Db 666 TCTTCCAGAGAGATCCCGCTCAACAACTTTCTTGTGAGCCACGCGCGCGGATCCGCG 725
QY 518 TGTTCACATCACTGCGACGTTTACAGAGGATTCGACTTGCACATTAATCAACCAAT 577
Db 726 TCTTCAATCTTACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 785
QY 578 TGGAGAGATTGAAGTCTGAAGTCACTCCGATACCTAGTTCCTTTGAGTCCAAAGAA 637
```



```

PRIOR APPLICATION NUMBER: 60/225,447
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/218,290
PRIOR FILING DATE: 2000-07-14
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 187
LENGTH: 378
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (347)
OTHER INFORMATION: n equals a,t,g, or c
US-10-242-515-187

```

```

Query Match          22.2%; Score 323.8; DB 16; Length 378;
Best Local Similarity 96.5%; Pred. No. 2.1e-97;
Matches 363; Conservative 0; Mismatches 8; Indels 5; Gaps 3;

```

```

QY 532 GCCAGCTCAGCAGGATCCCATCTGCACTACTACCACTTGGAGAGCATGAA 591
DB 1 GCCAGCTCAGCAGGATCCCATCTGCACTACTACCACTTGGAGAGCATGAA 60
QY 592 GTCTGAACTCCTCCGATACCTAGTTCTTTGCACTGCAAAACAGCTTGA 651
DB 61 GTCTGAACTCCTCCGATACCTAGTTCTTTGCACTGCAAAACAGCTTGA 120
QY 652 CTGCTCCGCTGGTGTATGATCACTGAGTGAAGCCCAACAGAGGAGCAT 711
DB 121 CTGCTCCGCTGGTGTATGATCACTGAGTGAAGCCCAACAGAGGAGCAT 180
QY 712 GTTCGCGAGCTGATGACTTACATGAGTGCATTCCTATGATGATGTTA 771
DB 181 GTTCGCGAGCTGATGACTTACATGAGTGCATTCCTATGATGATGTTA 240
QY 772 GACCTCCTCAGCAGCTGAAATCCAGCCTCTATGAGCCGATCTTATGAT 831
DB 241 GACCTCCTCAGCAGCTGAAATCCAGCCTCTATGAGCCGATCTTATGAT 300
QY 832 ATTGACAGTATTAATTATCTAGCTTTTGAAGATGCAATTTGATGATCA 891
DB 301 A-TGCACGATTAATTATCTAGCTTTTGAAGATGCAATTTGATGATCA 355
QY 892 GAGAGTCTCGAGGC 907
DB 356 GAGAGTCTCGAGGC 371

```

RESULT 12

```

US-10-467-595-50
Sequence 50, Application US/10467595
Publication No. US20040166501A1
GENERAL INFORMATION:
APPLICANT: AZIMZAI, Yalda; YUE, Henry;
APPLICANT: DING, Li; NGUYEN, Daniel B.;
APPLICANT: GANDHI, Ameena R.; BURFORD, Neil;
APPLICANT: THANGAVELU, Kavitha; ELLIOTT, Vicki S.;
APPLICANT: RAMKUMAR, Javalaxmi; YAO, Monique G.;
APPLICANT: LAU, Preeti G.; TANG, Y. Tom;
APPLICANT: SWARNAKAR, Anita; WAREEN, Bridget A.;
APPLICANT: WALIA, Nandinder K.; POLICKY, Jennifer L.;
APPLICANT: XU, Yuning; HONCHEN, Cynthia D.;
APPLICANT: AU-YOUNG, Janice K.; BAUGHN, Mariah R.;
APPLICANT: DUGGAN, Brendan M.; LU, Dying Aina M.;
APPLICANT: GIETZEN, Kimberly J.; JACKSON, Jennifer L.;
APPLICANT: RAUMANN, Bridget E.; LU, Yan;
APPLICANT: KAREHT, Stephanie K.; TRAN, Uyen K.;
APPLICANT: RICHARDSON, Thomas W.; EMBELING, Brook M.;
APPLICANT: HAFALIA April J.A.; BURRILL, John D.;
APPLICANT: MARCUS, Gregory A.; ZINGLER, Kurt A.;
APPLICANT: KABLE, Amy E.; GORVAD, Ann E.

```

```

TITLE OF INVENTION: RECEPTORS AND MEMBRANE-ASSOCIATED PROTEINS
FILE REFERENCE: PF-0894 PCT
CURRENT APPLICATION NUMBER: US/10/467,595
PRIOR FILING DATE: 2003-08-06
PRIOR APPLICATION NUMBER: PCT/US02/03868
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/267,201
PRIOR FILING DATE: 2001-02-06
PRIOR APPLICATION NUMBER: US 60/269,580
PRIOR FILING DATE: 2001-02-16
PRIOR APPLICATION NUMBER: US 60/282,679
PRIOR FILING DATE: 2001-04-09
PRIOR APPLICATION NUMBER: US 60/288,295
PRIOR FILING DATE: 2001-05-02
PRIOR APPLICATION NUMBER: US 60/348,687
PRIOR FILING DATE: 2002-01-14
NUMBER OF SEQ ID NOS: 86
SOFTWARE: PERL Program
SEQ ID NO 50
LENGTH: 1550
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Inocyte ID No: 195336CB1
US-10-467-595-50

```

```

Query Match          20.4%; Score 298; DB 17; Length 1550;
Best Local Similarity 100.0%; Pred. No. 2.6e-88;
Matches 298; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 101 AGCTATGCTTGTAGCTGGGGAAGTTGAAGGAAGGTTAAAGTCCAGTTG 160
DB 185 AGCTATGCTTGTAGCTGGGGAAGTTGAAGGAAGGTTAAAGTCCAGTTG 244
QY 161 ATGACATACCAAAATGAGGAGACCTACCTATTTATTTTAAAGAAAG 220
DB 245 ATGACATACCAAAATGAGGAGACCTACCTATTTATTTTAAAGAAAG 304
QY 221 GATTGACCTTCAACGAGAAAGAAATGGAAATGGACACTTCCCATTA 280
DB 305 GATTGACCTTCAACGAGAAAGAAATGGAAATGGACACTTCCCATTA 364
QY 281 GGTCCCGCTGACGGGAGAGCTGGAGGTTAGGCAATGTGAGAGATGCT 340
DB 365 GGTCCCGCTGACGGGAGAGCTGGAGGTTAGGCAATGTGAGAGATGCT 424
QY 341 TCACCATCAACCGGACCTTACCTCATCATCATGACCAAGATTCCTT 398
DB 425 TCACCATCAACCGGACCTTACCTCATCATCATGACCAAGATTCCTT 482

```

RESULT 13

```

US-09-844-948-3
Sequence 3, Application US/09844948
Publication No. US20030115161A1
GENERAL INFORMATION:
APPLICANT: Meyers, Rachel A.
APPLICANT: Williamson, Mark
TITLE OF INVENTION: 32132. A NOVEL FUCOSYLTRANSFERASE FAMILY
FILE REFERENCE: 10448-04801
CURRENT APPLICATION NUMBER: US/09/844,948
PRIOR FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: US 60/200,604
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 1479
TYPE: DNA
ORGANISM: Homo sapiens
US-09-844-948-3

```



```
FILE REFERENCE: MRI-006B
CURRENT APPLICATION NUMBER: US/09/814,353
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/191,031
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: US 60/207,124
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 60/211,940
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 60/216,820
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: US 60/220,661
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/257,672
NUMBER OF SEQ ID NOS: 22037
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9698
LENGTH: 495
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 368, 444
OTHER INFORMATION: n = A,T,C or G
US-09-814-353-9698
```

```
Query Match      30.8% Score 450.6; DB 10; Length 495;
Best Local Similarity 97.5%; Pred. No. 6.2e-140;
Matches 467; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
```

```
QY 663 GGTGTATGTACAGTCACTGTGACCCAGCATCAGACAGGAGAGAGCTATGTTGGCAGCT 722
DB 17 GCGCCAGGTACAGTCACTGTGACCCAGCATCAGACAGGAGAGAGCTATGTTGGCAGCT 76
QY 723 GATGCTTACATCGAGGTCGATTCCTATGATGATGTTTACGAAACAAAGACCTCCCTCA 782
DB 77 GATGCTTACATCGAGGTCGATTCCTATGATGATGTTTACGAAACAAAGACCTCCCTCA 136
QY 783 GCAGCTGAAAAATCCAGCTCTATGATGATGCGATGCGCTTTTATAGATCATTGCAAGTA 842
DB 137 GCAGCTGAAAAATCCAGCTCTATGATGATGCGATGCGCTTTTATAGATCATTGCAAGTA 196
QY 843 TAACTTTTCTAGCTTTTGAGAAATGCACTTTGATGATCATACTACAGAACTTCTG 902
DB 197 TAACTTTTCTAGCTTTTGAGAAATGCACTTTGATGATCATACTACAGAACTTCTG 256
QY 903 GAGGCCACTGAACTGGGGGT-AGTCCCTGATATTAATCCAGATCCCGAGCATCAGACT 961
DB 257 GAGGCCACTGAACTGGGGGTAACTCCCTGATATTAATCCAGATCCCGAGCATCAGACT 316
QY 962 GCGTCCCAAGTAACAAAAGTGTCTATTTCTGTATCAGAAATTTTCTCACCCGAGAACTG3 1021
DB 317 GCGTCCCAAGTAACAAAAGTGTCTATTTCTGTATCAGAAATTTTCTCACCCGAGAACTG3 376
QY 1022 CAAGTTAATACAGACGATGATCTGATGACATGTTAGGCGCTAGTGAATGTA 1081
DB 377 CAAGTTAATACAGACGATGATCTGATGACATGTTAGGCGCTAGTGAATGTA 1436
QY 1082 AGCTGAAGGATGAGATCTCTATACACGACGACTTTCAGACGCTCTCAGGGAACGAAATG 1140
DB 437 AGCTGAAGGATGAGATCTCTATACACGACGACTTTCAGACGCTCTCAGGGAACGAAATG 495
```

```
RESULT 10
US-09-764-877-187
Sequence 187, Application US/09764877
Patent No. US20020147140A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005
CURRENT APPLICATION NUMBER: US/09/764,877
```

```
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PAM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 187
LENGTH: 378
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (347)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-877-187
```

```
Query Match      22.2% Score 323.8; DB 9; Length 378;
Best Local Similarity 96.5%; Pred. No. 2.1e-97;
Matches 363; Conservative 0; Mismatches 8; Indels 5; Gaps 3;
```

```
QY 532 GCCACGTTGACAGGAGATCCCACTTGCCACTACTACCAATTAATCTTGAGAGCAATTGA 591
DB 1 GCCACGTTGACAGGAGATCCCACTTGCCACTACTACCAATTAATCTTGAGAGCAATTGA 60
QY 592 GTCTGAGTCACTCCGATACCTAGTTCCTTGACAGTCCCAAAACAGCTTGAAAAAGA 651
DB 61 GTCTGAGTCACTCCGATACCTAGTTCCTTGACAGTCCCAAAACAGCTTGAAAAAGA 120
QY 652 CTGCTCCGCTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 711
DB 121 CTGCTCCGCTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 712 GTTGGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 771
DB 181 GTTGGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 772 GACCTCCCTCAGCAGCTGAAAAATCCAGCTCTATGATGATGATGATGATGATGATGAT 831
DB 241 GACCTCCCTCAGCAGCTGAAAAATCCAGCTCTATGATGATGATGATGATGATGATGAT 300
QY 832 ATTGACAGTATATGTTATCTTCTAGCTTTTGAGATGACGTTTGATGATGATGATGAT 891
DB 301 ATTGACAGTATATGTTATCTTCTAGCTTTTGAGATGACGTTTGATGATGATGATGAT 355
QY 892 GAGAAGTTCTGAGGC 907
DB 356 GAGAAGTTCTGAGGC 371
```

```
RESULT 11
US-10-242-515-187
Sequence 187, Application US/10242515
Publication No. US20040009488A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005C1
CURRENT APPLICATION NUMBER: US/10/242,515
CURRENT FILING DATE: 2002-09-13
PRIOR APPLICATION NUMBER: 09/764,877
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 60/179,065
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: 60/214,886
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: 60/217,487
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,758
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,963
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/217,496
PRIOR FILING DATE: 2000-07-11
```

```
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 16082
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-16082

Query Match      31.0%; Score 453; DB 10; Length 530;
Best Local Similarity 98.9%; Pred. No. 1e-140;
Matches 456; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 663 GGTGTATGTACAGTCAGTGTGACCCACATCAGACAGGACAGTATGTCGGAGCT 722
DB 70 GCGCGAGGTACAGTCAGTGTGACCCACATCAGACAGGACAGTATGTCGGAGCT 129
QY 723 GATGACTTACATCAGAGTGCATTCCTATGATGTAATGTTTACGAAACAAAGACCTCCCTCA 782
DB 130 GATGACTTACATCAGAGTGCATTCCTATGATGTAATGTTTACGAAACAAAGACCTCCCTCA 189
QY 783 GCAGCTGAAATTCACAGCTCTATGATGCGATGCGATGCGCTTTTATAGATCATTCGACAGTA 842
DB 190 GCAGCTGAAATTCACAGCTCTATGATGCGATGCGATGCGCTTTTATAGATCATTCGACAGTA 249
QY 843 TAAGTTTATCCTAGCTTTTGAAGATGACATTTGTGATGATCACTACATGAGAACTTGTG 902
DB 250 TAAGTTTATCCTAGCTTTTGAAGATGACATTTGTGATGATCACTACATGAGAACTTGTG 309
QY 903 GAGGCCACTGAACTGGGGGTAGTCCCTGTATATTAACGATCCCCAGATCAGACACTG 962
DB 310 GAGGCCACTGAACTGGGGGTAGTCCCTGTATATTAACGATCCCCAGATCAGACACTG 369
QY 963 GCTTCCAAAGTAACAAAGTGTATCTTGTATCAGAAATTTCTCACCCAGGAACTGGC 1022
DB 370 GCTTCCAAAGTAACAAAGTGTATCTTGTATCAGAAATTTCTCACCCAGGAACTGGC 429
QY 1023 AAGTTACATCAGACGACTGGATTCGTATGACAGATTTGATGAGGCTATGTAGATGAA 1082
DB 430 AAGTTACATCAGACGACTGGATTCGTATGACAGATTTGATGAGGCTATGTAGATGAA 489
QY 1083 GCTGAGGCTGAGATCTCTAACACGAGCACTTCTGACAGCTC 1123
DB 490 GCTGAGGCTGAGATCTCTAACACGAGCACTTCTGACAGCTC 530

RESULT 8
US-09-814-353-3379
; Sequence 3379, Application US/09814353
; Publication No. US2003016581A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
```

```
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3379
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 368-444
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-3379

Query Match      30.8%; Score 450.6; DB 10; Length 495;
Best Local Similarity 97.5%; Pred. No. 6.2e-140;
Matches 467; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 663 GGTGTATGTACAGTCAGTGTGACCCACATCAGACAGGACAGTATGTCGGAGCT 722
DB 17 GCGCGAGGTACAGTCAGTGTGACCCACATCAGACAGGACAGTATGTCGGAGCT 76
QY 723 GATGACTTACATCAGAGTGCATTCCTATGATGTAATGTTTACGAAACAAAGACCTCCCTCA 782
DB 77 GATGACTTACATCAGAGTGCATTCCTATGATGTAATGTTTACGAAACAAAGACCTCCCTCA 136
QY 783 GCAGCTGAAATTCACAGCTCTATGATGCGATGCGATGCGCTTTTATAGATCATTCGACAGTA 842
DB 137 GCAGCTGAAATTCACAGCTCTATGATGCGATGCGATGCGCTTTTATAGATCATTCGACAGTA 196
QY 843 TAAGTTTATCCTAGCTTTTGAAGATGACATTTGTGATGATCACTACATGAGAACTTGTG 902
DB 197 TAAGTTTATCCTAGCTTTTGAAGATGACATTTGTGATGATCACTACATGAGAACTTGTG 256
QY 903 GAGGCCACTGAACTGGGGGTAGTCCCTGTATATTAACGATCCCCAGATCAGACACT 961
DB 257 GAGGCCACTGAACTGGGGGTAGTCCCTGTATATTAACGATCCCCAGATCAGACACT 316
QY 962 GCTTCCAAAGTAACAAAGTGTATCTTGTATCAGAAATTTTCTCACCCAGGAACTGG 1021
DB 317 GCTTCCAAAGTAACAAAGTGTATCTTGTATCAGAAATTTTCTCACCCAGGAACTGG 376
QY 1022 CAAGTTACATCAGACGACTGGATTCGTATGACAGATTTGATGAGGCTATGTAGATGAA 1081
DB 377 CAAGTTACATCAGACGACTGGATTCGTATGACAGATTTGATGAGGCTATGTAGATGAA 436
QY 1082 AGCTAAGGGGTGAGATCTCTAACACGAGCACTTCTGACAGCTCTCAGGAAACGAAATGG 1140
DB 437 AGCTAAGGGGTGAGATCTCTAACACGAGCACTTCTGACAGCTCTCAGGAAACGAAATGG 495

RESULT 9
US-09-814-353-9698
; Sequence 9698, Application US/09814353
; Publication No. US2003016583A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
```

```
QY 461 AACATTATTAAGCTCTTTTCATTAACGAGTATTAACCTTGTCACTACACTGCGACGTTTC 540
DB 2299 AACATTATTAAGCTCTTTTCATTAACGAGTATTAACCTTGTTCACACTACACTGCGACGTTTC 2240
QY 541 AGCAGGACCTCCCACTTGCCCACTACTAATCCCAATACCTTGGAGAGCATTTGAAGCCCGAG 600
DB 2239 AGCAGGACCTCCCACTTGCCCACTACTAATCCCAATACCTTGGAGAGCATTTGAAGCCCGAG 2180
QY 601 TCATCCGATTAACCTAGTTCCTTTGCGAGTCCCAAAAACAAGTTGAAAAAGACTTGTCTCG 660
DB 2179 TCATCCGATTAACCTAGTTCCTTTGCGAGTCCCAAAAACAAGTTGAAAAAGACTTGTCTCG 2120
QY 661 CTGGGTATATGATACGAGCTGTGACCCCAACATCGACGAGGAGCATATGTTGCGGAG 720
DB 2119 CTGGGTATATGATACGAGCTGTGACCCCAACATCGACGAGGAGCATATGTTGCGGAG 2060
QY 721 CTGATGACTTACATCGAGCTGTGATTCCTATGATGATGTTTACGAAAACAAGACCTCCCT 780
DB 2059 CTGATGACTTACATCGAGCTGTGATTCCTATGATGATGTTTACGAAAACAAGACCTCCCT 2000
QY 781 CAGCAGCTGAAAAATCCAGCTCTATGAGATGCGGATGCGCTTTATATGATCATTTGACAG 840
DB 1999 CAGCAGCTGAAAAATCCAGCTCTATGAGATGCGGATGCGCTTTATATGATCATTTGACAG 1940
QY 841 TATAAGTTTATCTAGCTTTTGAAGATGACAGTTTGTATGATCACTACCTGAGAAAGTTTC 900
DB 1939 TATAAGTTTATCTAGCTTTTGAAGATGACAGTTTGTATGATCACTACCTGAGAAAGTTTC 1880
QY 901 TGGAGGCGACCTGAAAACCTGGGGGTAGTCCCTGTATATTAACGATCCCCAGCATCAGAGAC 960
DB 1879 TGGAGGCGACCTGAAAACCTGGGGGTAGTCCCTGTATATTAACGATCCCCAGCATCAGAGAC 1820
QY 961 TGGCTTCCAAAGTAAACAAAGTGTATTTCTTGTATCAGAAATTTTCTGACCCGAGGAGAC 1020
DB 1819 TGGCTTCCAAAGTAAACAAAGTGTATTTCTTGTATCAGAAATTTTCTGACCCGAGGAGAC 1760
QY 1021 GCAAGTTACATCAGACGACTGATTTCTGATGACAGATTTGATGAGGCTATGATGATG 1080
DB 1759 GCAAGTTACATCAGACGACTGATTTCTGATGACAGATTTGATGAGGCTATGATGATG 1700
QY 1081 AAGCTGAAGGGGTAGATCTCTAACACGAGACTTGTGACAGCTCTCAGGGAGCGAAATGG 1140
DB 1699 AAGCTGAAGGGGTAGATCTCTAACACGAGACTTGTGACAGCTCTCAGGGAGCGAAATGG 1640
QY 1141 GAGAGTCAAGAGCTCAACCAAGACATTAATCATCGATGATTTGATGATGATGATGATG 1200
DB 1639 GAGAGTCAAGAGCTCAACCAAGACATTAATCATCGATGATTTGATGATGATGATGATG 1580
QY 1201 AAGCTGTGGGCTATATATGAGCTTCAAGAAAAGGGCTTACACCCAAAGATGGAGGCA 1260
DB 1579 AAGCTGTGGGCTATATATGAGCTTCAAGAAAAGGGCTTACACCCAAAGATGGAGGCA 1520
QY 1261 GAAGATACCCCACTGAGTGGCCAGAGGCCACAGTGTTCCTTCTCACAACCTCCGGAAT 1320
DB 1519 GAAGATACCCCACTGAGTGGCCAGAGGCCACAGTGTTCCTTCTCACAACCTCCGGAAT 1460
QY 1321 CCACCTTTGAGCTCTTTGCGAGAGATGTGATTTCCAGCTTTTGAACATCCAAAGAA 1380
DB 1459 CCACCTTTGAGCTCTTTGCGAGAGATGTGATTTCCAGCTTTTGAACATCCAAAGAA 1400
QY 1381 GCCCAGGACCTAAGGTGGCTGTTGATGAGATCAAAACTTTTCATCTCAGAGTTTGG 1440
DB 1399 GCCCAGGACCTAAGGTGGCTGTTGATGAGATCAAAACTTTTCATCTCAGAGTTTGG 1340
QY 1441 GGCCTAGATTTCAAGGACTGA 1461
DB 1339 GGCCTAGATTTCAAGGACTGA 1319
```

RESULT 6  
US-09-864-408A-7105  
; Sequence 7105, Application US/09864408A  
; Publication No. US20040009474A1

```
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Shimkets, Richard A.  
; TITLE OF INVENTION: No. US20040009474A1el Human polynucleotides and Polypeptides Enco  
; FILE REFERENCE: 21402-012  
; CURRENT APPLICATION NUMBER: US/09/864,408A  
; PRIOR FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: 60/206,690  
; NUMBER OF SEQ ID NOS: 9068  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7105  
; LENGTH: 597  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-864-408A-7105  
  
Query Match 39.8%; Score 581.2; DB 11; Length 597;  
Best Local Similarity 99.3%; Pred. No. 9.1e-184;  
Matches 594; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
  
QY 462 TCATGAAGATGTCCTCCGAAAAACAATTAAGCTCTTTTCATTAACGAGTATTAACCTTGT 521  
DB 1 TCATGAAGATGTCCTCCG-AAAAACATTAATTAAGCTCTTTTCATTAACGAGTATTAACCTTGT 59  
QY 522 CCACTACACTGCCAGCTTGACGAGGATTCCTCCACTTGCACCTAATCAATCACTTGTGA 581  
DB 60 CCACTACACTGCCAGCTTGACGAGGATTCCTCCACTTGCACCTAATCAATCACTTGTGA 119  
QY 582 GAGCATTGAAGTCTTGAAGTCACTCCGATCACTAGTTCCTTTGCGAGTCCAAACAAGCT 641  
DB 120 GAGCATTGAAGTCTTGAAGTCACTCCGATCACTAGTTCCTTTGCGAGTCCAAACAAGCT 179  
QY 642 TAGAAAAAGACTGCTCCGCTGGTGTATGATGACGATGACGATGACGATGACGATGACG 701  
DB 180 TAGAAAAAGACTGCTCCGCTGGTGTATGATGATGATGATGATGATGATGATGATGATG 239  
QY 702 GAGCAGCTATGTTGCGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 761  
DB 240 GAGCAGCTATGTTGCGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 299  
QY 762 ACGAAACAAAGACTCCCTCCAGAGCTGAAAAATCCAGCTCTATGAGAGCCGATGCTT 821  
DB 300 ACGAAACAAAGACTCCCTCCAGAGCTGAAAAATCCAGCTCTATGAGAGCCGATGCTT 359  
QY 822 TTATAGATCATTTGACAGATTAATGATTTATCTAGCTTTTGAAGATGACAGTTTGTATG 881  
DB 360 TTATAGATCATTTGACAGATTAATGATTTATCTAGCTTTTGAAGATGACAGTTTGTATG 419  
QY 882 CTATCATCTGAGAGTCTTGAGAGCCACTGAAACTGGGGGTAGTCCCTGTATTTACG 941  
DB 420 CTGATCATCTGAGAGTCTTGAGAGCCACTGAAACTGGGGGTAGTCCCTGTATTTACG 479  
QY 942 ATCCCCAGCATCAGACTGCTTCCAAAGTAAACAAAGTGTATTTCTGTATCAGAAAT 1001  
DB 480 ATCCCCAGCATCAGACTGCTTCCAAAGTAAACAAAGTGTATTTCTGTATCAGAAAT 539  
QY 1002 TTCTACCCCAAGGAACTGGCAAGTTATCATCAGACGATGATTTCTGATGACAGATTG 1059  
DB 540 TTCTACCCCAAGGAACTGGCAAGTTATCATCAGACGATGATTTCTGATGACAGATTG 597
```

RESULT 7  
US-09-814-353-16082  
; Sequence 16082, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lilie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER

Page 5

```

US-09-814-353-20757/c
; Sequence 20757, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Illlie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20757
; LENGTH: 2937
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20757

Query Match
Matches Local Similarity 97.8%; Score 1428.6; DB 10; Length 2937;
Matches 1448; Conservative 0; Mismatches 9; Indels 4; Gaps 1;

QY 1 ATGAAGTCACGGGGCTCTCCCGACGAGTTACAGACTCCATGCATGCTTCAATGATCAG 60
Db 2775 ATGAAGTCACGGGGCTCTCCCGACGAGTTACAGACTCCATGCATGCTTCAATGATCAG 2716

QY 61 TGGCCCTTATCTCAACACGAGGAGGAGGACACATMAAAGAGTCATGGTGGAGTGGG 120
Db 2715 TGGCCCTTATCTCAACACGAGGAGGAGGACACATMAAAGAGTCATGGTGGAGTGGG 2656

QY 121 AAGTTGAAGAGAGATTMAAAAGTTCCAGTTTGCAAGATGACATACAAAAATGAG 180
Db 2715 AAGTTGAAGAGAGATTMAAAAGTTCCAGTTTGCAAGATGACATACAAAAATGAG 2596

QY 2655 AAGTTGAAGAGAGATTMAAAAGTTCCAGTTTGCAAGATGACATACAAAAATGAG 2596
Db 181 GAAGACCTCAGCATTCTTAATTCATTTCTTAAGAAAGAGATTACCTTCAACGAGAA 240

QY 2595 GAAGACCTCAGCATTCTTAATTCATTTCTTAAGAAAGAGATTACCTTCAACGAGGT 2536
Db 241 AGMAAATGGGAATTGACAGCTACCCCATTTGCTGTGTGTCCCGCTGACGGGGAG 300

QY 2535 GGA----TCGCGAGTGACAGCTACCCCATTTGCTGTGTGTCCCGCTGACGGGGAG 2480
Db 301 ACTGGAGATTAGGCTCATGTGAGAGCATGCTTTTCTTACCATCAACCGACCTTAC 360

QY 2479 ACTGGAGATTAGGCTCATGTGAGAGCATGCTTTTCTTACCATCAACCGACCTTAC 2420
Db 361 CTCGATCATCATGACATGACCAAGCATTCCTTTCTATGTAAGTACGCTTTAAGATGATAGC 420

QY 2419 CTCGATCATCATGACATGACCAAGCATTCCTTTCTATGTAAGTACGCTTTAAGATGATAGC 2360
Db 421 TTACCTCTGCTCGGAAAGCCCATATGATGAGGCTGTTTTTCATGAAGAATCCCCGAAA 480

QY 2359 TTACCTCTGCTCGGAAAGCCCATATGATGAGGCTGTTTTTCATGAAGAATCCCCGAAA 2300

```

Db 739 ACTGGAGGTTAGGCCAATGTGAGCAGATGCTTTGTTCTTCCACATCAACCGACTTAC 798  
Qy 361 CTCATCATCATGACGACCAAGCATTCCTCTCTATGTACTGACTTTAACAATAGATAC 420  
Db 739 CTCATCATCATGACCAAGCATTCCTCTCTATGTACTGACTTTAACAATAGATAC 858  
Qy 421 TTACCTGCTGGGAAAGCCATCATGATGGGCTGTTTTCATATAAGTCCCGAAA 480  
Db 859 TTACCTGCTGGGAAAGCCATCATGATGGGCTGTTTTCATATAAGTCCCGAAA 918  
Qy 481 AACATTAAGGCTCTTCAATTAACCAAGTATTAACCTTGTCACTACATGCGCATTC 540  
Db 919 AACATTAAGGCTCTTCAATTAACCAAGTATTAACCTTGTCACTACATGCGCATTC 978  
Qy 541 AGCAGGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCT 600  
Db 979 AGCAGGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCT 1038  
Qy 601 TCACCTCCGATACCTGATTCCTTTCGATCCCAAAAACAAGTTCGAAAGCTTCGCG 660  
Db 1039 TCACCTCCGATACCTGATTCCTTTCGATCCCAAAAACAAGTTCGAAAGCTTCGCG 1098  
Qy 661 CTGCTGTAATGACGTCAGACTGTCAGCCACCATCAGACGGGACGCTATGTTGCGAG 720  
Db 1099 CTGCTGTAATGACGTCAGACTGTCAGCCACCATCAGACGGGACGCTATGTTGCGAG 1158  
Qy 721 CTGATGACTTACATCGAGTGCATTCCTATGATGATGATGATGATGATGATGATGAT 780  
Db 1159 CTGATGACTTACATCGAGTGCATTCCTATGATGATGATGATGATGATGATGATGAT 1218  
Qy 781 CAGAGCTGAAAAATCCAGCCTCTATGATGATGATGATGATGATGATGATGATGAT 840  
Db 1219 CAGAGCTGAAAAATCCAGCCTCTATGATGATGATGATGATGATGATGATGATGAT 1278  
Qy 841 TATAGTTTATCCAGCTTTTGAATGAGCTTTGATGATGATGATGATGATGATGATGAT 900  
Db 1279 TATAGTTTATCCAGCTTTTGAATGAGCTTTGATGATGATGATGATGATGATGATGAT 1338  
Qy 901 TGAAGGCTAGAACTGGGGGTAGTCCCTGTATATTAACGATCCCGACATCAAGAC 960  
Db 1339 TGAAGGCTAGAACTGGGGGTAGTCCCTGTATATTAACGATCCCGACATCAAGAC 1398  
Qy 961 TGGCTTCCAACTAACAAGTCTATTTCTGTATCAGAAATTTTCTACCCCGAGAACTG 1020  
Db 1399 TGGCTTCCAACTAACAAGTCTATTTCTGTATCAGAAATTTTCTACCCCGAGAACTG 1458  
Qy 1021 GCAAGTTACATCAGACGCTGATTTCTGATGACGATTTGATGAGGCTATGATGATGAT 1080  
Db 1459 GCAAGTTACATCAGACGCTGATTTCTGATGACGATTTGATGAGGCTATGATGATGAT 1518  
Qy 1081 AAGCTGAAGGCTGATTTCTTAACCAAGCCTTCTGACAGCTTCAAGGAAACGAAATGG 1140  
Db 1519 AAGCTGAAGGCT -AGATCTCTAACCAAGCCTTCTGACAGCTTCAAGGAAACGAAATGG 1577  
Qy 1141 GAGAGCAAGACGTCACACAGGACATTAATCATGATGATGATGATGATGATGATGATGAT 1200  
Db 1578 GAGAGCAAGACGTCACACAGGACATTAATCATGATGATGATGATGATGATGATGATGAT 1637  
Qy 1201 AAGGTGTGGGTAAATATCAGGCTTCAAGAAAAGGCTTACACCCAAAGATGGAGGCA 1260  
Db 1638 AAGGTGTGGGTAAATATCAGGCTTCAAGAAAAGGCTTACACCCAAAGATGGAGGCA 1697  
Qy 1261 GAAAGTACCCAGCTGATTTGCGGACAGCCACAGTGTGCTTCTTCAACCTCCGAGCT 1320  
Db 1698 GAAAGTACCCAGCTGATTTGCGGACAGCCACAGTGTGCTTCTTCAACCTCCGAGCT 1757  
Qy 1321 CCACCTTTCAGCTCTTTCGAGAGATGAGATTTCCAGCTTGAACATCCAGAAAGAA 1380  
Db 1758 CCACCTTTCAGCTCTTTCGAGAGATGAGATTTCCAGCTTGAACATCCAGAAAGAA 1817  
Qy 1381 GCCCAGGCACTAAGGTGCTGTTGATAGGATTAACATTTTCATCTCAAGGTTTGG 1440  
Db 1818 GCCCAGGCACTAAGGTGCTGTTGATAGGATTAACATTTTCATCTCAAGGTTTGG 1877

Qy 1441 GCCCTAGATATTCAGAGCTGA 1461  
Db 1878 GGCCTAGATATTCAGAGCTGA 1898  
RESULT 4  
US-09-876-997-43  
; Sequence 43/Application US/09876997  
; Publication No. US20030152921A1  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, Jean Baptiste  
; APPLICANT: Bouquelere, Lydie  
; APPLICANT: Jobert, Severin  
; TITLE OR INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS  
; FILE REFERENCE: 78.US4.CIP  
; CURRENT APPLICATION NUMBER: US/09/876, 997  
; PRIOR FILING DATE: 2001-06-08  
; PRIOR APPLICATION NUMBER: US 09/731, 872  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/187, 470  
; PRIOR FILING DATE: 2000-03-06  
; PRIOR APPLICATION NUMBER: US 60/169, 629  
; PRIOR FILING DATE: 1999-12-08  
; NUMBER OF SEQ ID NOS: 482  
; SOFTWARE: Patent.pm  
; SEQ ID NO 43  
; LENGTH: 1942  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 334..1551  
; NAME/KEY: sig peptide  
; LOCATION: 334..426  
; OTHER INFORMATION: Von Heijne matrix  
; OTHER INFORMATION: score 4.0554926521937  
; OTHER INFORMATION: seq TVEFLVTLQALDI/VE  
; US-09-876-997-43  
Query Match 99.0%; Score 1445.8; DB 10; Length 1942;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1458; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
Qy 1 ATGAAGTCAAGGAGGCTTCCGAGGAGTTACAGACTCCATCAATGCTTCATATATGAG 60  
Db 439 ATGAAGTCAAGGAGGCTTCCGAGGAGTTACAGACTCCATCAATGCTTCATATATGAG 498  
Qy 61 TGGCTTATCTAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120  
Db 499 TGGCTTATCTAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 558  
Qy 121 AAGTTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 180  
Db 559 AAGTTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 618  
Qy 181 GAAAGCCTACGATCTTAATTTCTTAAGAAAGAGGATTAACCTTCAACAGGAAA 240  
Db 619 GAAAGCCTACGATCTTAATTTCTTAAGAAAGAGGATTAACCTTCAACAGGAAA 678  
Qy 241 AGAAATGGGAATTTGAGACGCTACCCATTATGCTCTGCTGCTCCCGGTGACGGGGAG 300  
Db 679 AGAAATGGGAATTTGAGACGCTACCCATTATGCTCTGCTGCTCCCGGTGACGGGGAG 738  
Qy 301 ACTGGAGGATTAGGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360  
Db 739 ACTGGAGGATTAGGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 798  
Qy 361 CTCATCATCATGACGACCAAGCATTCCTCTCTATGTACTGACTTTAACAATAGATAC 420  
Db 799 CTCATCATCATGACGACCAAGCATTCCTCTCTATGTACTGACTTTAACAATAGATAC 858  
Qy 421 TTACCTGCTGGGAAAGCCATCATGATGGGCTGTTTTCATATAAGTCCCGAAA 480

Db 343 GAAGACCTACGATCTTAATTTATTTTAAAGAAAGATTACCTTCAACAGGAAA 402  
Qy 241 AGAAATGGGAATTGACAGCTAACCCTATTAATGCTTGTGTGCTCCCTGACGAGGAG 300  
Db 403 AGAAATGGGAATTGACAGCTAACCCTATTAATGCTTGTGTGCTCCCTGACGAGGAG 462  
Qy 301 ACAGGAGGTTAGGCAATGTGGAGAGATGCTGTTTTCATCAGCATCAACGACCTAC 360  
Db 463 ACTGGAGGTTAGGCAATGTGGAGAGATGCTGTTTTCATCAGCATCAACGACCTAC 522  
Qy 361 CTCATCATCAGATGACCAAGAGATTCCTTTCTATGATGATGATGATGATGATGATG 420  
Db 523 CTCATCATCAGATGACCAAGAGATTCCTTTCTATGATGATGATGATGATGATGATG 582  
Qy 421 TTACCTCTGCTGGAAAGCCATGATGATGATGATGATGATGATGATGATGATGATG 480  
Db 583 TTACCTCTGCTGGAAAGCCATGATGATGATGATGATGATGATGATGATGATGATG 642  
Qy 481 AACAATTATAGCTCTTTCAATAAACAGATGATGATGATGATGATGATGATGATGATG 540  
Db 643 AACAATTATAGCTCTTTCAATAAACAGATGATGATGATGATGATGATGATGATGATG 702  
Qy 541 AGAAGCATTCCTACCTGCTACCTACCTACCTACCTACCTACCTACCTACCTACCTAC 600  
Db 703 AGAAGCATTCCTACCTGCTACCTACCTACCTACCTACCTACCTACCTACCTACCTAC 762  
Qy 601 TCATCTGCTGCTGGAAAGCCATGATGATGATGATGATGATGATGATGATGATGATG 660  
Db 763 TCATCTGCTGCTGGAAAGCCATGATGATGATGATGATGATGATGATGATGATGATG 822  
Qy 661 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720  
Db 823 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 882  
Qy 721 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780  
Db 883 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 942  
Qy 781 CAGCAGCTGAAATCAGCCTCTATGATGATGATGATGATGATGATGATGATGATGATG 840  
Db 943 CAGCAGCTGAAATCAGCCTCTATGATGATGATGATGATGATGATGATGATGATGATG 1002  
Qy 841 TATAAGTTTATCTAGCTTTTGAATGATGATGATGATGATGATGATGATGATGATG 900  
Db 1003 TATAAGTTTATCTAGCTTTTGAATGATGATGATGATGATGATGATGATGATGATG 1062  
Qy 901 TGAAGGCACTGAAATGAGGAGTATGATGATGATGATGATGATGATGATGATGATG 960  
Db 1063 TGAAGGCACTGAAATGAGGAGTATGATGATGATGATGATGATGATGATGATGATG 1122  
Qy 961 TGGCTTCAAGTAAACAAAGTCTATTTTATGATGATGATGATGATGATGATGATG 1020  
Db 1123 TGGCTTCAAGTAAACAAAGTCTATTTTATGATGATGATGATGATGATGATGATG 1182  
Qy 1021 GCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
Db 1183 GCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1242  
Qy 1081 AAGCTAAGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
Db 1243 AAGCTAAGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1302  
Qy 1141 GAGAGTCAAGAGTCAACAGAGATTAATGATGATGATGATGATGATGATGATGATG 1200  
Db 1303 GAGAGTCAAGAGTCAACAGAGATTAATGATGATGATGATGATGATGATGATGATG 1362  
Qy 1201 AAGGTGAGGCTAATATCAGGCTTGAAGAAAGGCTTACCAACCAAAAGATGGAGCA 1260  
Db 1363 AAGGTGAGGCTAATATCAGGCTTGAAGAAAGGCTTACCAACCAAAAGATGGAGCA 1422  
Qy 1261 GAAAGTACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320

Db 1423 GAAGATACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1482  
Qy 1321 CCACCTTTAGCTCTTTTGGAGAGATGATGATGATGATGATGATGATGATGATGATG 1380  
Db 1483 CCACCTTTAGCTCTTTTGGAGAGATGATGATGATGATGATGATGATGATGATGATG 1542  
Qy 1381 GCCCAGGACTAAGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1440  
Db 1543 GCCCAGGACTAAGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1602  
Qy 1441 GGCCTAGTATTCAGGACTGA 1461  
Db 1603 GGCCTAGTATTCAGGACTGA 1623

RESULT 3  
US-09-731-872-43  
Sequence 43, Application US/09731872  
Patent No. US20020102604A1  
GENERAL INFORMATION:  
APPLICANT: Dumas Mline Edwards, Jean Baptiste  
APPLICANT: Bouquelere, Lydie  
APPLICANT: Jober, Severin  
TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS  
FILE REFERENCE: 78. US. REG  
CURRENT APPLICATION NUMBER: US/09/731.872  
CURRENT FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/169,629  
PRIOR FILING DATE: 1999-12-08  
PRIOR APPLICATION NUMBER: US 60/187,470  
PRIOR FILING DATE: 2000-03-06  
NUMBER OF SEQ ID NOS: 482  
SOFTWARE: Patent.pm  
SEQ ID NO 43  
LENGTH: 1942  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 334..1551  
NAME/KEY: sig peptide  
LOCATION: 334..426  
OTHER INFORMATION: Von Heijne matrix  
OTHER INFORMATION: score 4.0554925521937  
OTHER INFORMATION: seq IVFLVLTQLADLV/VE  
US-09-731-872-43

Query Match  
Best Local Similarity 99.0%; Score 1445.8; DB 9; Length 1942;  
Matches 1458; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 ATGAAGTACAGGAGCTTCCCGAGGAGTTACAGATCTCATGCAATCTTCAATGATGAG 60  
Db 439 ATGAAGTACAGGAGCTTCCCGAGGAGTTACAGATCTCATGCAATCTTCAATGATGAG 498  
Qy 61 TGGCTTATATCAACACAGAGAGAGAGAGATTAAGAGATGATGATGATGATGATGATG 120  
Db 499 TGGCTTATATCAACACAGAGAGAGAGATTAAGAGATGATGATGATGATGATGATGATG 558  
Qy 121 AAGTTTGAAGAGAGATTTTAAAGTCCAGTTTGCAGATGAGATACAAATAATGAG 180  
Db 559 AAGTTTGAAGAGAGATTTTAAAGTCCAGTTTGCAGATGAGATACAAATAATGAG 618  
Qy 181 GAAGCACTAGCATCTTAAATTTCTTAAAGAAAGAGATGATGATGATGATGATGATG 240  
Db 619 GAAGCACTAGCATCTTAAATTTCTTAAAGAAAGAGATGATGATGATGATGATGATG 678  
Qy 241 AGAAATAGGAAATGAGACAGTACCCCATTAATGCTGTGTGTGCTGACGAGGAGAG 300  
Db 679 AGAAATAGGAAATGAGACAGTACCCCATTAATGCTGTGTGTGCTGACGAGGAGAG 758  
Qy 301 ACTGGAGGTTAAGGCAATGATGAGAGATGATGATGATGATGATGATGATGATGATGATG 360

|    |      |                 |                       |                   |                            |      |
|----|------|-----------------|-----------------------|-------------------|----------------------------|------|
| QY | 6    | TGGCCCTTTATCTAA | CACCAAGACAGCGAGACA    | CATTAAGAAGAGTCA   | TGCTGATGCTGGG              | 120  |
| Db | 61   | TGGCCTTTATCTAA  | CACCAAGACAGCGAGACA    | CATTAAGAAGAGTCA   | TGCTGATGCTGGG              | 120  |
| QY | 121  | AAGTTGAAGAAGAGG | ATTAAAGTTTCCAGTTTG    | CAGATGACATACAAAAT | TGAG                       | 180  |
| Db | 121  | AAGTTGAAGAAGAGG | ATTAAAGTTTCCAGTTTG    | CAGATGACATACAAAAT | TGAG                       | 180  |
| QY | 181  | GAAGCACTACAGCAT | CTTAATTCATTTCTTA      | AGAAAGAGGATGAG    | CCCTTCAACAGGAA             | 240  |
| Db | 181  | GAAGCACTACAGCAT | CTTAATTTCTTAAGAAAGAGG | ATGAGCTTCAACAGGAA |                            | 240  |
| QY | 241  | AGAAAATGGGAAT   | TGAGACGTACCCCATTA     | TGCTGTGTGTC       | CCCCGTGACGGGGAG            | 300  |
| Db | 241  | AGAAAATGGGAAT   | TGAGACGTACCCCATTA     | TGCTGTGTGTC       | CCCCGTGACGGGGAG            | 300  |
| QY | 301  | ACTGGGAGGTTAG   | GGCCATGTGAGACAGAT     | TCCTTTTCA         | CCATCAACGGACCTAC           | 360  |
| Db | 301  | ACTGGGAGGTTAG   | GGCCATGTGAGACAGAT     | TCCTTTTCA         | CCATCAACGGACCTAC           | 360  |
| QY | 361  | CTCCATCATCATAC  | TACCAAGACATTCCTCT     | CTATGTA           | CTAGCTTTACATATATAGC        | 420  |
| Db | 361  | CTCCATCATCATAC  | TACCAAGACATTCCTCT     | CTATGTA           | CTAGCTTTACATATATAGC        | 420  |
| QY | 421  | TTTACTCTGCTCG   | GAAGCCCATCATGACTGGG   | CTGTTTTTCATG      | AGAGTCCCGGAA               | 480  |
| Db | 421  | TTTACTCTGCTCG   | GAAGCCCATCATGACTGGG   | CTGTTTTTCATG      | AGAGTCCCGGAA               | 480  |
| QY | 481  | AACAATTAATTA    | CTCTTTTCA             | TAAACCAAGTATAC    | CTGTGTCATACACCTGACAGTTC    | 540  |
| Db | 481  | AACAATTAATTA    | CTCTTTTCA             | TAAACCAAGTATAC    | CTGTGTCATACACCTGACAGTTC    | 540  |
| QY | 541  | AGCAGGCAT       | TCCCATCTGCCACTA       | CTAATCCCAAT       | CTTGAGAGCATTAAGTCTGTAG     | 600  |
| Db | 541  | AGCAGGCAT       | TCCCATCTGCCACTA       | CTAATCCCAAT       | CTTGAGAGCATTAAGTCTGTAG     | 600  |
| QY | 601  | TCATCCCGAT      | TACCTGATCTCTTG        | CGAGTCCMAAAACA    | AGTTTGAAAGAACTTGTCTCG      | 660  |
| Db | 601  | TCATCCCGAT      | TACCTGATCTCTTG        | CGAGTCCMAAAACA    | AGTTTGAAAGAACTTGTCTCG      | 660  |
| QY | 661  | CTGGTATATGTA    | CTAGCTAGACTGTGAC      | CCACATCA          | CAGAGGACAGCTATGTTGGCAG     | 720  |
| Db | 661  | CTGGTATATGTA    | CTAGCTAGACTGTGAC      | CCACATCA          | CAGAGGACAGCTATGTTGGCAG     | 720  |
| QY | 721  | CTGATGACTTA     | CATGAGAGTGGAT         | TCTCTATGTGTA      | AGTTTACGAAACAACCTCCCT      | 780  |
| Db | 721  | CTGATGACTTA     | CATGAGAGTGGAT         | TCTCTATGTGTA      | AGTTTACGAAACAACCTCCCT      | 780  |
| QY | 781  | CACGACCTA       | AAAAATCCAGCTCTAT      | TGATGCCAT         | TGGCTTTATATGATCATTCACAG    | 840  |
| Db | 781  | CACGACCTA       | AAAAATCCAGCTCTAT      | TGATGCCAT         | TGGCTTTATATGATCATTCACAG    | 840  |
| QY | 841  | TATTAAGTTTAT    | CTAGCTTTGAGATG        | CACTTGTGAT        | CACTACATCACTGAGAAATTG      | 900  |
| Db | 841  | TATTAAGTTTAT    | CTAGCTTTGAGATG        | CACTTGTGAT        | CACTACATCACTGAGAAATTG      | 900  |
| QY | 901  | TGAGAGCC        | CTAATACTGGGGT         | AGTCCCTG          | TATATAGCATCCCGCAGATCA      | 960  |
| Db | 901  | TGAGAGCC        | CTAATACTGGGGT         | AGTCCCTG          | TATATAGCATCCCGCAGATCA      | 960  |
| QY | 961  | TGCTTTC         | CAAGTAACAAAGT         | CTATTTCTG         | TATCAAGATTTTCTAC           | 1020 |
| Db | 961  | TGCTTTC         | CAAGTAACAAAGT         | CTATTTCTG         | TATCAAGATTTTCTAC           | 1020 |
| QY | 1021 | GCAAGTTA        | CATCAGACGACTGG        | ATTCTGATG         | CAAGTTGATGAGGCTTATGTAGATGG | 1080 |
| Db | 1021 | GCAAGTTA        | CATCAGACGACTGG        | ATTCTGATG         | CAAGTTGATGAGGCTTATGTAGATGG | 1080 |
| QY | 1081 | AACCTAAAGGT     | GAGATCTTAAC           | CAGGAACTTGTG      | ACACTCTCAGGGAGCGAAATGG     | 1140 |
| Db | 1081 | AACCTAAAGGT     | GAGATCTTAAC           | CAGGAACTTGTG      | ACACTCTCAGGGAGCGAAATGG     | 1140 |
| QY | 1141 | GGAGTCA         | AGAGCTCAAC            | CAGGACATTA        | TCGATCATTTGATG             | 1200 |

|    |      |  |      |
|----|------|--|------|
| Db | 1141 | GGAGTGCAGAGACGTCAACCGAGACCAATTATCATCGATGCATTGGAGTGTGGTGACC   | 1200 |
| Qy | 1201 | AAGGCTGGGGCTAAATATGAGCTTCAGAGAAAAGGGCTTACACCCCAAAAGATGGAGGCA | 1260 |
| Db | 1201 | AAGGTGTGGCTAAATATGAGCTTCAGAGAAAAGGGCTTACACCCCAAAAGATGGAGGCA  | 1260 |
| Qy | 1261 | GAAAGTACCACCTGAGTTGGCCCAAGCCCAKAGTGTGTGCTTCTCACACTCCGGACT    | 1320 |
| Db | 1261 | GAAAGTACCACCTGAGTTGGCCCAAGCCCAKAGTGTGTGCTTCTCACACTCCGGACT    | 1320 |
| Qy | 1321 | CCACCTTTGAGCTCTTTCGAGAAATGTGATTTCCAGCTTTGCAACAAATCCAAGAAAGA  | 1380 |
| Db | 1321 | CCACCTTTGAGCTCTTTCGAGAAATGTGATTTCCAGCTTTGCAACAAATCCAAGAAAGA  | 1380 |
| Qy | 1381 | GCCCGAGGCACTAAGGTGGCTGTTGATGAGAAATCAAACTTTCACTCAAGAGTTTGG    | 1440 |
| Db | 1381 | GCCCGAGGCACTAAGGTGGCTGTTGATGAGAAATCAAACTTTCACTCAAGAGTTTGG    | 1440 |
| Qy | 1441 | GGCTTAGTATTTAAGGACTGA  | 1461 |
| Db | 1441 | GGCTTAGTATTTAAGGACTGA  | 1461 |

```

RESULT 2
US-10-080-960-1
, Sequence 1, Application US/10080960
, Publication No. US20020197695A1
, GENERAL INFORMATION:
, APPLICANT: Millennium Pharmaceuticals, Inc.
, APPLICANT: Glucksmann, Maria
, APPLICANT: Meyers, Rachel
, TITLE OF INVENTION: 80090, 52874, 52880, 63497, AND 33425
, TITLE OF INVENTION: METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES THEREOF
, FILE REFERENCE: 38155-2004.00
, CURRENT APPLICATION NUMBER: US/10/080,960
, CURRENT FILING DATE: 2001-10-19
, PRIOR APPLICATION NUMBER: US 60/242,040
, PRIOR FILING DATE: 2000-10-20
, PRIOR APPLICATION NUMBER: US 60/242,038
, PRIOR FILING DATE: 2000-10-20
, PRIOR APPLICATION NUMBER: US 60/241,992
, PRIOR FILING DATE: 2000-10-20
, PRIOR APPLICATION NUMBER: US 60/242,637
, PRIOR FILING DATE: 2000-10-23
, NUMBER OF SEQ. ID NOS: 37
, SOFTWARE: FastsEQ for Windows Version 4.0
, SEQ ID NO 1
, LENGTH: 1669
, TYPE: DNA
, ORGANISM: Homo sapiens
, FEATURE:
, NAME/KEY: CDS
, LOCATION: (163)...(1623)
US-10-080-960-1

```

|    | Query Match           | 100.0%   | Score 1461  | DB 14      | Length 1669 |
|----|-----------------------|--|-------------|------------|-------------|
|    | Best Local Similarity | 100.0%   | Prod. No. 0 |            |             |
|    | Matches 1461          | Conservative   | 0           | Mismatches | 0           |
|    |                       |  |             | Indels     | 0           |
|    |                       |  |             | Gaps       | 0           |
| QY | 1                     | ATGAAAGTCACGGGGCCCTCCCCAGGAGATTACAGATCTCATGCAATGCTTCATGATCAG | 60          |            |             |
| DB | 163                   | ATGAAAGTCACGGGGCCCTCCCCAGGAGATTACAGATCTCATGCAATGCTTCATGATCAG | 222         |            |             |
| QY | 61                    | TGCGCTTTATCTTAACCCAGAGCAGAGACATTAAGAGTCTATGGTTGACCTGGGG      | 120         |            |             |
| DB | 223                   | TGCGCTTTATCTTAACCCAGAGCAGAGACATTAAGAGTCTATGGTTGACCTGGGG      | 282         |            |             |
| QY | 121                   | AAGTTGGAAGAGAGAGATTAAAGTCTCCAGTTTCGAATGAGATCAAAAGATGAG       | 180         |            |             |
| DB | 283                   | AAGTTGGAAGAGAGAGATTAAAGTCTCCAGTTTCGAATGAGATCAAAAGATGAG       | 342         |            |             |
| QY | 181                   | GAAGCACTAGCATCTTAATTCATTCTTTAGAAAGAGATTAAGCTTCAACAGGAAA      | 240         |            |             |

| Result No. | Score  | Query Match | length | DB | ID                  | Description           |
|------------|--------|-------------|--------|----|---------------------|-----------------------|
| 1          | 1461   | 100.0       | 1461   | 14 | US-10-080-960-3     | Sequence 3, Appl1     |
| 2          | 1461   | 100.0       | 1669   | 14 | US-10-080-960-1     | Sequence 1, Appl1     |
| 3          | 1445.8 | 99.0        | 1942   | 9  | US-09-731-872-43    | Sequence 43, Appl1    |
| 4          | 1445.8 | 99.0        | 1942   | 10 | US-09-876-997-43    | Sequence 43, Appl1    |
| 5          | 1428.6 | 97.8        | 2937   | 10 | US-09-814-353-20757 | Sequence 20757, Appl1 |
| 6          | 581.2  | 39.8        | 597    | 11 | US-09-864-608A-7105 | Sequence 7105, Appl1  |
| 7          | 453    | 31.0        | 530    | 10 | US-09-814-353-16082 | Sequence 16082, Appl1 |
| 8          | 450.6  | 30.8        | 495    | 10 | US-09-814-353-3379  | Sequence 3379, Appl1  |
| 9          | 450.6  | 30.8        | 495    | 10 | US-09-814-353-3698  | Sequence 3698, Appl1  |
| 10         | 323.8  | 22.2        | 378    | 9  | US-09-764-877-187   | Sequence 187, Appl1   |
| 11         | 323.8  | 22.2        | 378    | 15 | US-10-242-515-187   | Sequence 187, Appl1   |
| 12         | 298    | 20.4        | 1550   | 17 | US-10-467-595-50    | Sequence 50, Appl1    |
| 13         | 180.8  | 12.4        | 1479   | 10 | US-09-844-948-3     | Sequence 3, Appl1     |
| 14         | 180.8  | 12.4        | 1479   | 13 | US-10-184-648-21    | Sequence 21, Appl1    |

|    |       |      |       |    |                      |                        |
|----|-------|------|-------|----|----------------------|------------------------|
| 15 | 180.8 | 12.4 | 2557  | 10 | US-09-884-948-1      | Sequence 1, Appl1      |
| 16 | 180.8 | 12.4 | 2557  | 13 | US-10-194-698-9      | Sequence 19, Appl1     |
| 17 | 180.8 | 12.4 | 2566  | 10 | US-09-814-353-19076  | Sequence 19076, Appl1  |
| 18 | 166   | 11.4 | 2408  | 16 | US-10-108-2804-730   | Sequence 730, Appl1    |
| 19 | 150.8 | 10.3 | 1388  | 15 | US-10-072-977-13     | Sequence 13, Appl1     |
| 20 | 113.4 | 7.8  | 500   | 10 | US-09-814-353-12805  | Sequence 12805, Appl1  |
| 21 | 107.8 | 7.4  | 477   | 10 | US-09-818-995-32851  | Sequence 2051, Appl1   |
| 22 | 102.4 | 7.0  | 502   | 10 | US-09-814-353-422    | Sequence 22, Appl1     |
| 23 | 102.4 | 7.0  | 502   | 10 | US-09-814-353-4420   | Sequence 6420, Appl1   |
| 24 | 52.6  | 3.6  | 1586  | 17 | US-10-437-963-5998   | Sequence 5998, Appl1   |
| 25 | 45.4  | 3.1  | 954   | 13 | US-10-442-559-123113 | Sequence 123113, Appl1 |
| 26 | 39.8  | 2.7  | 1650  | 13 | US-10-425-114-30550  | Sequence 30550, Appl1  |
| 27 | 39.8  | 2.7  | 1813  | 13 | US-10-425-114-13954  | Sequence 13950, Appl1  |
| 28 | 38.8  | 2.7  | 1371  | 17 | US-10-437-963-10850  | Sequence 10850, Appl1  |
| 29 | 38.6  | 2.6  | 376   | 17 | US-10-437-963-44572  | Sequence 44572, Appl1  |
| 30 | 38.2  | 2.6  | 596   | 17 | US-10-021-233-1208   | Sequence 7208, Appl1   |
| 31 | 37.2  | 2.5  | 7498  | 15 | US-10-311-453-529    | Sequence 229, Appl1    |
| 32 | 36.6  | 2.5  | 3729  | 16 | US-10-369-493-55316  | Sequence 25316, Appl1  |
| 33 | 36.4  | 2.5  | 31703 | 16 | US-10-088-117-1712   | Sequence 172, Appl1    |
| 34 | 36.2  | 2.5  | 3584  | 9  | US-09-784-077-1      | Sequence 1, Appl1      |
| 35 | 36.2  | 2.5  | 3584  | 13 | US-10-700-505-1      | Sequence 1, Appl1      |
| 36 | 35.4  | 2.4  | 1660  | 15 | US-10-355-975-2      | Sequence 2, Appl1      |
| 37 | 35.2  | 2.4  | 567   | 13 | US-10-424-559-55350  | Sequence 55350, Appl1  |
| 38 | 35.2  | 2.4  | 11745 | 15 | US-10-240-453-506    | Sequence 206, Appl1    |
| 39 | 35    | 2.4  | 241   | 9  | US-09-604-2874-401   | Sequence 401, Appl1    |
| 40 | 35    | 2.4  | 241   | 10 | US-09-551-621-401    | Sequence 401, Appl1    |
| 41 | 35    | 2.4  | 241   | 14 | US-10-007-805-401    | Sequence 401, Appl1    |
| 42 | 35    | 2.4  | 241   | 15 | US-10-076-632-401    | Sequence 401, Appl1    |
| 43 | 35    | 2.4  | 241   | 15 | US-10-124-885-401    | Sequence 401, Appl1    |
| 44 | 35    | 2.4  | 386   | 15 | US-10-060-036-8854   | Sequence 2854, Appl1   |
| 45 | 35    | 2.4  | 2857  | 9  | US-09-923-302-88     | Sequence 88, Appl1     |

```

RESULT 1
US-10-080-960-3
; Sequence 3, Application US/10080960
; Publication No. US20020197695A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Glucksmann, Maria
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 800990, 52874, 52880, 63497, AND 33425
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES THEREOF
; FILE REFERENCE: 38155-20044.00
; CURRENT APPLICATION NUMBER: US/10/080,960
; CURRENT FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/242,040
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/242,038
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/241,992
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/242,637
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-080-960-3

Query Match          100.0%; Score 1461; DB 14; Length 1461;
Best Local Similarity 100.0%; Pred No. 0;
Matches 1461; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATAAAGTACGGGGCCCTCCAGGAGTTACAGACTCCAGCAATGTTTAATGATCAG 60

Db 1 ATCAAGTACGGGGCCCTCCAGGAGTTACAGACTCCAGCAATGTTCAATGATCAG 60



**This Page Blank (uspto)**

Mon Sep 13 09:12:39 2004

us-10-080-960-3.rml

Page 7

Job time : 80.8182 secs

---

QY 484 AATTATAGCT 494  
DB 81 AATCCAGCT 71

RESULT 14  
US-08-155-331-12  
; Sequence 12, Application US/08155331  
; Patent No. 5441931  
; GENERAL INFORMATION:  
; APPLICANT: Foster, Donald C  
; APPLICANT: Sprecher, Cindy  
; APPLICANT: No. 5441931, KJeld  
; TITLE OF INVENTION: NOVEL HUMAN AMYLOID PROTEIN PRECURSOR  
; TITLE OF INVENTION: HOMOLOG AND KUNITZ-TYPE INHIBITOR  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Zymogenetics, Inc.  
; STREET: 4225 Roosevelt Way, N.E.  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/155,331  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,692  
; FILING DATE: 02-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, Gary E  
; REGISTRATION NUMBER: 31-684  
; REFERENCE/DOCKET NUMBER: 92-21C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-547-8080 ext 322  
; TELEFAX: 206-548-2329  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3725 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: APFH  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 73..2364  
US-08-155-331-12

Query Match 2.4%; Score 35; DB 1; Length 3725;  
Best Local Similarity 54.2%; Pred. No. 0.4;  
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 364 CATCATCATGACCAAGCATTCCTTCTATGTAAGTACTTACATAGATAGCTTA 423  
DB 1588 CCTATGCAATTCCTCCAGGCTTATAGCGCTTATGTCGAGTACGAAAGATGCTTA 1587  
QY 424 CCTGTGCTCGGAAAGCCCATCATGACTGAGTGGCTGTTTTCATGAAAGTCCCGGAAAAAC 483  
DB 1588 CATACCATCCGTCATTCACGACATGTGTGCTGTGACCCGAAAGGCGGCCAGATG 1647  
QY 484 AATTATAGCT 494  
DB 1648 AATCCAGCT 1658

RESULT 15  
US-08-424-022-12  
; Sequence 12, Application US/08424022  
; Patent No. 5677146  
; GENERAL INFORMATION:  
; APPLICANT: Foster, Donald C  
; APPLICANT: Sprecher, Cindy  
; APPLICANT: No. 5677146, KJeld  
; TITLE OF INVENTION: NOVEL HUMAN AMYLOID PROTEIN PRECURSOR  
; TITLE OF INVENTION: HOMOLOG AND KUNITZ-TYPE INHIBITOR  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Zymogenetics, Inc.  
; STREET: 4225 Roosevelt Way, N.E.  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/424,022  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,692  
; FILING DATE: 02-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, Gary E  
; REGISTRATION NUMBER: 31-684  
; REFERENCE/DOCKET NUMBER: 92-21C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-547-8080 ext 322  
; TELEFAX: 206-548-2329  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3725 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: APFH  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 73..2364  
US-08-424-022-12

Query Match 2.4%; Score 35; DB 1; Length 3725;  
Best Local Similarity 54.2%; Pred. No. 0.4;  
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 364 CATCATCATGACCAAGCATTCCTTCTATGTAAGTACTTACATAGATAGCTTA 423  
DB 1588 CCTATGCAATTCCTCCAGGCTTATAGCGCTTATGTCGAGTACGAAAGATGCTTA 1587  
QY 424 CCTGTGCTCGGAAAGCCCATCATGACTGAGTGGCTGTTTTCATGAAAGTCCCGGAAAAAC 483  
DB 1588 CATACCATCCGTCATTCACGACATGTGTGCTGTGACCCGAAAGGCGGCCAGATG 1647  
QY 484 AATTATAGCT 494  
DB 1648 AATCCAGCT 1658

Search completed: September 13, 2004, 02:21:05

Best Local Similarity 54.2%; Pred. No. 0.04;  
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 364 CATCATCATGACCAAGCAATTCCTCTTATGTAGTACTTAAATAGATAGCTTA 423  
DB 201 CCTCATCGCATTCCTCCAGGCTTACGCGGTATGTCGCTGAGAAACAAAGATCGCTTA 142  
QY 424 CCTGTGCTCGGAAAGCCCATCATGACTGGCTGTTTTCATGAAGAGTCCCGAAAAAC 483  
DB 141 CATACCATCGGTATATACAGCATGTGTGCTGTGACCCAGAAAGCGGCCCATGATG 82  
QY 484 AATTATAGCT 494  
DB 81 AATCCAGGT 71

RESULT 11  
US-09-433-826B-401/c  
Sequence 401, Application US/09433826B

GENERAL INFORMATION:  
PATENT NO. 6579973  
APPLICANT: Jiang, Yugu  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Xu, Jiangchun  
APPLICANT: Harlocker, Susan L.  
TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE  
FILE REFERENCE: 210121.470C4  
CURRENT APPLICATION NUMBER: US/09/433,826B  
CURRENT FILING DATE: 1999-11-03  
NUMBER OF SEQ ID NOS: 474  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 401  
LENGTH: 241  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(241)  
OTHER INFORMATION: n = A,T,C or G  
US-09-433-826B-401

Query Match 2.4%; Score 35; DB 4; Length 241;  
Best Local Similarity 54.2%; Pred. No. 0.04;  
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 364 CATCATCATGACCAAGCAATTCCTCTTATGTAGTACTTAAATAGATAGCTTA 423  
DB 201 CCTCATCGCATTCCTCCAGGCTTACGCGGTATGTCGCTGAGAAACAAAGATCGCTTA 142  
QY 424 CCTGTGCTCGGAAAGCCCATCATGACTGGCTGTTTTCATGAAGAGTCCCGAAAAAC 483  
DB 141 CATACCATCGGTATATACAGCATGTGTGCTGTGACCCAGAAAGCGGCCCATGATG 82  
QY 484 AATTATAGCT 494  
DB 81 AATCCAGGT 71

RESULT 12

US-09-604-287A-401/c  
Sequence 401, Application US/09604287A

GENERAL INFORMATION:  
PATENT NO. 6586572  
APPLICANT: Jiang, Yugu  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Xu, Jiangchun  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Hepler, William T.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER

FILE REFERENCE: 210121.470C7  
CURRENT APPLICATION NUMBER: US/09/604,287A  
CURRENT FILING DATE: 2000-06-22  
NUMBER OF SEQ ID NOS: 489  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 401  
LENGTH: 241  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(241)  
OTHER INFORMATION: n = A,T,C or G  
US-09-604-287A-401

Query Match 2.4%; Score 35; DB 4; Length 241;  
Best Local Similarity 54.2%; Pred. No. 0.04;  
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 364 CATCATCATGACCAAGCAATTCCTCTTATGTAGTACTTAAATAGATAGCTTA 423  
DB 201 CCTCATCGCATTCCTCCAGGCTTACGCGGTATGTCGCTGAGAAACAAAGATCGCTTA 142  
QY 424 CCTGTGCTCGGAAAGCCCATCATGACTGGCTGTTTTCATGAAGAGTCCCGAAAAAC 483  
DB 141 CATACCATCGGTATATACAGCATGTGTGCTGTGACCCAGAAAGCGGCCCATGATG 82  
QY 484 AATTATAGCT 494  
DB 81 AATCCAGGT 71

RESULT 13  
US-09-834-759-401/c  
Sequence 401, Application US/09834759

GENERAL INFORMATION:  
PATENT NO. 6680197  
APPLICANT: Jiang, Yugu  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Xu, Jiangchun  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Hepler, William T.  
APPLICANT: Henderson, Robert A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER  
FILE REFERENCE: 210121.470C9  
CURRENT APPLICATION NUMBER: US/09/834,759  
CURRENT FILING DATE: 2001-04-13  
NUMBER OF SEQ ID NOS: 547  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 401  
LENGTH: 241  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(241)  
OTHER INFORMATION: n = A,T,C or G  
US-09-834-759-401

Query Match 2.4%; Score 35; DB 4; Length 241;  
Best Local Similarity 54.2%; Pred. No. 0.04;  
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 364 CATCATCATGACCAAGCAATTCCTCTTATGTAGTACTTAAATAGATAGCTTA 423  
DB 201 CCTCATCGCATTCCTCCAGGCTTACGCGGTATGTCGCTGAGAAACAAAGATCGCTTA 142  
QY 424 CCTGTGCTCGGAAAGCCCATCATGACTGGCTGTTTTCATGAAGAGTCCCGAAAAAC 483  
DB 141 CATACCATCGGTATATACAGCATGTGTGCTGTGACCCAGAAAGCGGCCCATGATG 82





```
APPLICANT: DORNER, F.
APPLICANT: SCHERFLINGER, F.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)683-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZpc-Fls
US-08-232-463-14

Query Match
Best Local Similarity 9.5%; Score 44; DB 1; Length 7218;
Matches 29; Conservative 151; Mismatches 126; Indels 0; Gaps 0;

QY 79 AGAGCAGCGAGCATAAAGAGGTCATGCTGAGCTGGGAGATTGAAGGAGAG 138
DB 1308 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1249
QY 139 TTTAAAGTTCAGTTGTCAGATGACATACAAAATGGAGAGACCATGCTT 198
DB 1248 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1189
QY 199 AATCATTTCTTAGAAGAAGATGACTTCAACGAGAAAGAAATGGGAATTGAC 258
DB 1188 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1129
QY 259 AGTACCCCATTAATGCTCTGATGCTCCCGCTGACGAGGAGAGCTGGAGTTCAG 318
DB 1128 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1069
QY 319 TGTGAGACAGATGTTGTTTCTTACCATCAACCGACTCACTCATATACATGACC 378
DB 1068 RRATGCAAGCTCCCTGACCTGACGCAAGCTGGAATTAATTTGTGAGGATAGCA 1009
QY 379 AAAGCA 384
DB 1008 AAGCA 1003
```

```
RESULT 3
US-09-621-976-2813
Sequence 2813, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 2813
LENGTH: 832
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 235..399
US-09-621-976-2813

Query Match
Best Local Similarity 10.1%; Score 41.4; DB 4; Length 832;
Matches 36; Conservative 165; Mismatches 156; Indels 0; Gaps 0;

QY 896 AGTTCGAGGCGCACTGAAGTGGGGTGCCTGTATATTACGATCCCCAGCATCA 955
DB 13 AKCMTKMSMSYMMWYKMYKTYRMRKKKXAMKWKMTWYRMYRMYKXKKA 72
QY 956 CAGACTGGCTTCCAGTAACAAAGTGTATTCTGTATCAGAAATTTCTCACCCAGG 1015
DB 73 MCRKTKKKKKKKGYMMWYMGWRSSYMAWMTRTWTGYAYRSMYMWYRCWKKAAYRX 132
QY 1016 AACTGCAAGTTACATCAGACGAGTGGATTCGATGACAGATGTATAGAGGCTATAG 1075
DB 133 TTCTSSGKMTWYMRKAWTWTWKKTYAATRYMMWCTKWRASWYICMMWGKAK 192
QY 1076 AATGAGCTGAAGGGTGAATCTCTAACGACGACTTCTGACAGCTCTCAGGAAACGA 1135
DB 193 WSTWRSRSYASARSARCCYSCSWGMSWKYMMWRBGMWRTGAKMWRASCMERX 252
QY 1136 AATGGGAGTGCAGACGTCACACGAGCAATTACATGATGATTCATTTAGTGTATG 1195
DB 253 YAKSKTSYSKMMWCTRSWKYCYTARWTGYCYRKGMMWRGRMYASKKYMWRWW 312
QY 1196 GCACCAAGTGTGGCTAATATCAGGCTTCAAGAAAGGCTTACCACCAAAAGAT 1252
DB 313 CWRMYRYSIGTASWWRWYWTMMKWKYAMAPAAWRWMAWMAWRRCALAAAT 369

RESULT 4
US-09-621-976-18033
Sequence 18033, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 18033
LENGTH: 474
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2004, 21:46:05 / Search time 79.8182 Seconds  
(without alignments)  
10157.873 Million cell updates/sec

Title: US-10-080-960-3

Perfect score: 1461  
Sequence: 1 atgaagtcacggccccc.....gcctagatcgaagactga 1461

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/prodata/2/ina/5A.COMB.seq: \*  
2: /cgn2\_6/prodata/2/ina/5B.COMB.seq: \*  
3: /cgn2\_6/prodata/2/ina/5A.COMB.seq: \*  
4: /cgn2\_6/prodata/2/ina/5B.COMB.seq: \*  
5: /cgn2\_6/prodata/2/ina/PCTUS.COMB.seq: \*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Query Length | DB ID | Description         |
|------------|-------|-------------|--------------|-------|---------------------|
| 1          | 230.2 | 15.8        | 450          | 4     | US-09-621-976-12473 |
| 2          | 44    | 3.0         | 7218         | 1     | US-08-232-463-14    |
| 3          | 41.4  | 2.8         | 832          | 4     | US-09-621-976-2813  |
| 4          | 36.4  | 2.5         | 474          | 4     | US-09-621-976-18033 |
| 5          | 36.2  | 2.5         | 1814         | 2     | US-08-483-151-1     |
| 6          | 36.2  | 2.5         | 1814         | 5     | PCT-US96-06427-1    |
| 7          | 36    | 2.5         | 430          | 4     | US-09-621-976-16656 |
| 8          | 35.4  | 2.4         | 1060         | 4     | US-09-579-664B-2    |
| 9          | 35    | 2.4         | 241          | 4     | US-09-389-681-401   |
| 10         | 35    | 2.4         | 241          | 4     | US-09-620-405B-401  |
| 11         | 35    | 2.4         | 241          | 4     | US-09-433-825B-401  |
| 12         | 35    | 2.4         | 241          | 4     | US-09-604-287A-401  |
| 13         | 35    | 2.4         | 241          | 4     | US-09-834-759-401   |
| 14         | 35    | 2.4         | 3725         | 1     | US-08-155-331-12    |
| 15         | 35    | 2.4         | 3725         | 1     | US-08-424-022-12    |
| 16         | 35    | 2.4         | 3725         | 1     | US-08-424-017B-12   |
| 17         | 35    | 2.4         | 3725         | 5     | PCT-US93-11696-12   |
| 18         | 34.2  | 2.3         | 1086         | 1     | US-07-914-281-12    |
| 19         | 34.2  | 2.3         | 1086         | 1     | US-08-393-246-12    |
| 20         | 34.2  | 2.3         | 1086         | 1     | US-08-525-058A-12   |
| 21         | 34.2  | 2.3         | 1086         | 1     | US-08-696-731-12    |
| 22         | 34.2  | 2.3         | 1086         | 3     | US-09-042-531-12    |
| 23         | 34.2  | 2.3         | 2042         | 3     | US-09-063-237-2     |
| 24         | 34.2  | 2.3         | 2043         | 1     | US-07-914-281-1     |
| 25         | 34.2  | 2.3         | 2043         | 1     | US-08-393-246-1     |
| 26         | 34.2  | 2.3         | 2043         | 1     | US-08-273-411-4     |
| 27         | 34.2  | 2.3         | 2043         | 1     | US-08-525-058A-1    |

|    |      |     |        |   |                     |                   |
|----|------|-----|--------|---|---------------------|-------------------|
| 28 | 34.2 | 2.3 | 2043   | 2 | US-08-696-731-1     | Sequence 1, Appl1 |
| 29 | 34.2 | 2.3 | 2043   | 3 | US-09-042-531-1     | Sequence 1, Appl1 |
| 30 | 34.2 | 2.3 | 2043   | 5 | PCT-US91-00899-1    | Sequence 1, Appl1 |
| 31 | 34   | 2.3 | 7218   | 1 | US-08-232-463-14    | Sequence 14, Appl |
| 32 | 33.4 | 2.3 | 4892   | 4 | US-09-620-312D-1065 | Sequence 1065, Ap |
| 33 | 33.4 | 2.3 | 100848 | 4 | US-09-596-002-39    | Sequence 39, Appl |
| 34 | 32.6 | 2.2 | 1316   | 1 | US-07-914-281-10    | Sequence 10, Appl |
| 35 | 32.6 | 2.2 | 1316   | 1 | US-08-393-246-10    | Sequence 10, Appl |
| 36 | 32.6 | 2.2 | 1316   | 1 | US-08-525-058A-10   | Sequence 10, Appl |
| 37 | 32.6 | 2.2 | 1316   | 2 | US-08-696-731-10    | Sequence 10, Appl |
| 38 | 32.6 | 2.2 | 1316   | 3 | US-09-042-531-10    | Sequence 10, Appl |
| 39 | 32.6 | 2.2 | 1654   | 1 | US-07-914-281-13    | Sequence 13, Appl |
| 40 | 32.6 | 2.2 | 1654   | 1 | US-08-393-246-13    | Sequence 13, Appl |
| 41 | 32.6 | 2.2 | 1654   | 1 | US-08-525-058A-13   | Sequence 13, Appl |
| 42 | 32.6 | 2.2 | 1654   | 2 | US-08-696-731-13    | Sequence 13, Appl |
| 43 | 32.6 | 2.2 | 1654   | 2 | US-09-042-531-13    | Sequence 13, Appl |
| 44 | 32.4 | 2.2 | 1527   | 4 | US-09-244-111-7     | Sequence 7, Appl1 |
| 45 | 32.4 | 2.2 | 3035   | 2 | US-08-723-624-18    | Sequence 18, Appl |

## ALIGNMENTS

```

RESULT 1
US-09-621-976-12473
Sequence 12473, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Maline Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTS and Encoded Human Proteins.
FILE REFERENCE: GENSET. 054PR2
CURRENT APPLICATION NUMBER: US/09/621, 976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 12473
LENGTH: 450
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-12473

Query Match      15.8%      Score 230.2; DB 4; Length 450;
Best Local Similarity 96.7%; Pred. No. 9.1e-73;
Matches 235; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      1219 AGGCTTCAGAAAGGGCTTACCAACCCAAAGATGGAGGAGAGATACCACTGAGT 1278
      |||||
Db      71  AGGCAAGGAATAGGGGCTTACCAACCCAAAGATGGAGGAGAGATACCACTGAGT 130

QY      1279 TGGCCAGAGCCCAAGTGTGTTGCTTCTCACCACCTCCGAGATCCACTTTGAGGCTTTG 1338
      |||||
Db      131 TGGCCAGAGCCCAAGTGTGTTGCTTCTCACCACCTCCGAGATCCACTTTGAGGCTTTG 190

QY      1339 CCAGAGATGTGATTTCCAGCTTGAACATCCAGAAAGAGCCGAGCACTAAGTGG 1398
      |||||
Db      191 CCAGAGATGTGATTTCCAGCTTGAACATCCAGAAAGAGCCGAGCACTAAGTGG 250

QY      1399 CTGGTTGATAGGAATCAAACTTTCATCTCAAGAGTTTGGGGCTTAGATTCAAGAC 1458
      |||||
Db      251 CTGGTTGATAGGAATCAAACTTTCATCTCAAGAGTTTGGGGCTTAGATTCAAGAC 310

QY      1459 TGA 1461
      |||||
Db      311 TGA 313

RESULT 2
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:

```



gene; ss.  
 Homo sapiens.  
 MO200299056-A2.  
 12-DEC-2002.  
 05-JUN-2002; 2002MO-US017699.  
 06-JUN-2001; 2001US-0285869P.  
 11-JUL-2001; 2001US-0304121P.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Ruben SM, Bell A, Birse CE, Komatsoulis G, Choi GH, Olsen H;  
 Ni J, Baker KP;  
 MPI: 2003-140609/13.  
 P-FSD3; ABR54303.  
 Novel human secreted polypeptides and polynucleotides for diagnosing and treating neural, immune system, muscular, reproductive, gastrointestinal, cardiovascular, renal, proliferative disorders and cancerous diseases.  
 Claim 1; Page 551-552; 603pp; English.  
 The present invention relates to novel human secreted proteins (ABR54278-ABR54311) and their coding sequences (ACC62475-ACC62528). The proteins and their coding sequences are useful for treating, preventing, diagnosing and/or prognosing neoplastic diseases, blood disorders, cardiovascular disorders (e.g. cardiomyopathy, hypertension, hypotension), respiratory disorders (e.g. lung cancer, pneumonia, bronchitis, asthma), endocrine disorders (e.g. Addison's disease, Cushing's syndrome, hyperpituitarism, pituitary dwarfism, acromegaly, thyrotoxicosis), reproductive system disorders (e.g. premenstrual syndrome, polycystic ovary syndrome), infectious diseases caused by bacteria, fungal, viral, parasitic, protozoal, and/or blood-related disorders and infections, leukopenia, leukaemias, arthritis, asthma, autoimmune diseases, rheumatoid arthritis, immune deficiency, psoriasis, haemophilia, diabetes mellitus, allergies, and bone cancers, Paget's disease, gout, osteoporosis, arrhythmia, angina, prostate cancer, renal disorders, urolithiasis, Alzheimer's disease, Parkinson's disease, schizophrenia, attention deficit disorder, obsessive compulsive pneumonia, obesity, golfer, ulcerative colitis, hepatitis. They are also useful for stimulating epithelial cell proliferation and basal keratinocytes for wound healing, and to stimulate hair follicle production and healing of dermal wounds

Sequence 1804 BP; 405 A; 473 C; 456 G; 465 T; 0 U; 5 Other;

Query Match 11.1%; Score 162.2; DB 7; Length 1804;  
 Best Local Similarity 52.4%; Pred. No. 2e-41;  
 Matches 409; Conservative 1; Mismatches 359; Indels 12; Gaps 2;

QY 441 CCATCATGACTGGGCTGTTTTCATGAAGATCCCGAAACAAATTATAGTCTTTCA 500  
 DB 158 CCGTCGGGCTGGGCGCTCCACGAGAGTGGCCCTCAACAATCTTGTGTAGCCA 217  
 QY 501 TAAACCAAGTATGTTCTTTCACTACACTGCGAGTTTACAGGATTTCCACTTGGC 560  
 DB 218 CGGCCCGGCACTCGCCTCTTCAATCTTACCTCACCCTTAGTGCCACTCGATTACC 277  
 QY 561 ACTACTACCAATACATTGAGAGCATTTGAAGTCTGAAGTCACTCGATACCTAGTTCC 620  
 DB 278 GCTGTGGTGAAGGCTGCGCCGAGACCGCTATCTGCGC--CGCCGCTGCTCCGCC 334  
 QY 621 TTTCAGTCCAAAAACAAGCTTAGAAAAAGACTTCTCGCTGGTGTATGTAAGTCAGA 680  
 DB 335 CATGAACGCGCGGAGTGGCGCGCCGCGCTACGCGCCGCTCTATCTGACAGTACA 394  
 QY 681 CTGTGACCAACATCATGACAGGAGAGCTATGTTGGAGCTGATGACTTACATGAGGT 740

DB 395 CTGGAGTGCACGCGGACCGGACCGCTACGTGGCGGAGCTCATGCGCCACATCCCGGT 454  
 QY 741 CGATTCTATGATGATGTTTACGAAACAAAGCTTCCT-----CAGCACTGAA 791  
 DB 455 AGACTCTACGGAATGCTGCGATGCGAGTGGAGCTGCTTACCGCGGCTACAGACAC 514  
 QY 792 AAATCCAGCTCTATGATGATGCGATGCTTTTATAGATCATGTGACAGTATTAATTAT 851  
 DB 515 AGCAGCGCCACACCGAGGATCCAGAGCTTGGGTTTCTTGTCCCGCTATTAATTCCA 574  
 QY 852 CTTAGCTTTGAATGACAGTTTGTATGATCAATCACTGAGAACTTTGGAGGCCACT 911  
 DB 575 CTGGCCCTGGAATAAGTCCATCTGTACACATGACAGAAAACTGTGCGTCCCAT 634  
 QY 912 GAAATGCGGGGTAGTCCCTGTATATTACGATCCCGACATCAGACAGTGGCTTCCAA 971  
 DB 635 GCACCTGGGCGGTGCGCCGCTGACCGCGTCTCCTCTGTGAGGAGCTGATGCCAA 694  
 QY 972 TAACAAAAGTGTATTTCTGTATCAGAAATTTCTCAACCCAGGAACTGGCAAGTTAAT 1031  
 DB 695 CAATCACTCCGCTCATCTGATGATGATTTGAGTCTCTCAGAGCTGGCAGATTAT 754  
 QY 1032 CAGACAGTGGATTTGATGACAGATTTGATGAGGCTTATGAGATGGAAGCTGAAGG 1091  
 DB 755 TGACTTTCTGAACAAGATGATGAGATGTAATGAAATDCTTGCATACACAACTGG 814  
 QY 1152 CGTACACAGGAATTTACATGATGATTTAGTGTATGTTGTGACCAAGTGTGGC 1211  
 DB 875 TCCTTTGCTGCTTAACTACTTAAAGGCTTGAAGTGTTCGTGTGACTAGAACTGGC 934  
 QY 1212 T 1212  
 DB 935 T 935

Search completed: September 12, 2004, 22:22:34  
 Job time : 415.627 secs

KM hemostatic; antidiabetic; antiallergic; antigout; antiarrhythmic;  
 KM antianginal; nephrotoxic; neuroleptic; neuroprotective; noctropic;  
 KM antiparkinsonian; tranquilizer; anorectic; antitumor; hepatotoxic;  
 KM dermatological; gene therapy; neoplastic disease; blood disorder;  
 KM cardiovascular disorder; respiratory disorder; cancer;  
 KM endocrine disorder; reproductive system disorder; infection; arthritis;  
 KM asthma; autoimmune disease; immune deficiency; psoriasis;  
 KM diabetes mellitus; allergy; Paget's disease; gout; osteoporosis;  
 KM renal disorder; Alzheimer's disease; Parkinson's disease; schizophrenia;  
 KM attention deficit disorder; obsessive compulsive disorder; wound healing;  
 gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200299066-A2.  
 XX  
 PD 12-DEC-2002.  
 XX  
 PF 05-JUN-2002; 2002WO-US017699.  
 XX  
 PR 06-JUN-2001; 2001US-0295869P.  
 PR 11-JUL-2001; 2001US-0304121P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Bell A, Birse CE, Komatsculis G, Choi GH, Olsen H;  
 PI N1 J, Baker KP;  
 XX  
 DR WPI; 2003-140609/13.  
 DR P-FSDB; ABR54302.  
 XX  
 PT Novel human secreted polypeptides and polynucleotides for diagnosing and  
 PT treating neural, immune system, muscular, reproductive, gastrointestinal,  
 PT cardiovascular, renal, proliferative disorders and cancerous diseases.  
 XX  
 PS Claim 1; Page 551; 605bp; English.  
 XX  
 CC The present invention relates to novel human secreted proteins (ABR54278-  
 CC ABR54331) and their coding sequences (ACC62475-ACC62528). The proteins  
 CC and their coding sequences are useful for treating, preventing,  
 CC diagnosing and/or prognosing neoplastic diseases, blood disorders,  
 CC cardiovascular disorders (e.g. cardiomyopathy, hypertension,  
 CC bronchodilation, respiratory disorders (e.g. lung cancer, pneumonia,  
 CC hypotension), endocrine disorders (e.g. Addison's disease,  
 CC Cushing's syndrome, hyperpituitarism, pituitary dwarfism, acromegaly,  
 CC thyroiditis), reproductive system disorders (e.g. premenstrual syndrome,  
 CC polycystic ovary syndrome), infectious diseases caused by bacteria,  
 CC fungal, viral, parasitic, protozoal, and/or blood-related disorders and  
 CC infections, leukopenia, leukaemia, arthritis, asthma, autoimmune  
 CC diseases, rheumatoid arthritis, immune deficiency, psoriasis,  
 CC haemophilia, diabetes mellitus, allergies, and bone cancers, Paget's  
 CC disease, gout, osteoporosis, arthralgia, and prostate cancer, renal  
 CC disorders, urolithiasis, Alzheimer's disease, Parkinson's disease,  
 CC schizophrenia, attention deficit disorder, obsessive compulsive  
 CC pneumonia, obesity, goiter, ulcerative colitis, hepatitis. They are also  
 CC useful for stimulating epithelial cell proliferation and basal  
 CC keratinocytes for wound healing, and to stimulate hair follicle  
 CC production and healing of dermal wounds  
 XX  
 SQ Sequence 1812 BP; 431 A; 480 C; 461 G; 440 T; 0 U; 0 Other;  
 XX  
 Query Match 11.1%; Score 162.6; DB 7; Length 1812;  
 Best Local Similarity 52.8%; Pred. No. 1.5e-41;  
 Matches 410; Conservative 0; Mismatches 359; Indels 12; Gaps 2;  
 XX  
 QY 441 CCATCATGATGGGGCTTTTTCATGAGAGTCCCGAAGAAACATTTAAGCTCTTCA 500  
 Db 176 CCTGGCGCTGGGGCTCTCTCCACAGAGAGTGGCCCTCAACATCTTCTGAGAGCA 235  
 QY 501 TAAACGATGATTAATCTTGTTCATCACTGACAGCTTCCAGAGGCTTCCACTGGCC 560  
 Db 236 CGGCGCGGCGATCGGCTCTTCAATCTTCACTTCACTTCACTGAGTACCTGAGATACCC 295

QY 561 ACTAATACCCAAATCTTGGAGAGATTGAAGTCCGTAAGTCACTCCGATACCTAGTCC 620  
 Db 296 GCTGTGCTGCGAGTGGCTGCCCCGAGACCGCTATCTGGC---CGCCGCTGCTCCGCC 352  
 QY 621 TTTCAGATCCAAAACAAAGCTTAAAGAAAAGACTTCTCCGTGTGTATGATGACAGCA 680  
 Db 353 CATGAGAGCGCGGAGTGGCGCCCGCGGCTGACGCGCGCTCTATCTGACAGTACA 412  
 QY 681 CTGTGACCCACCATATAGACAGGACAGTATGTTCCGAGTGTGACTATACATGAGGT 740  
 Db 413 CTGCACGATGCGAGGACCGGAGCCGCTAGTGCAGGACTCAATGCGCACATCCGGT 472  
 QY 741 CGATTCCTATGGTGAATGTTTACGAAACAAAGACCTCCCT-----CAGCAGCTGAA 791  
 Db 473 AGACTCTTACGGAATGCTCTGAGAAATCGGAGACTGCTACCGCGGCTGACAGGAC 532  
 QY 792 AATTCAGCTCTATGATGATCCGATGCTTTTATGAGATCATTCACAGATATAGTTAT 851  
 Db 533 ACCCAGGCGCACACCGAGAGATCCAGAGCTCTTGGCTTTCTGTCGCGCTATAGTCCA 592  
 QY 852 CCTAGCTTTTGAAGATGAGTTCCTGATGATGATCATCATGAGAAATTCGAGGCGCACT 911  
 Db 593 CTGGCCCTGGAATATGCAATCTGTACAGATATGATGAGAAATGCTGGCTCCCAT 652  
 QY 912 GAACTGGGGGTAGTCCCTGTATATTAAGATCCCGCAGATCATCAGACTGGCTTCCAG 971  
 Db 653 GCACCTGGGCGGTGGCCCGTGTACCGGGTTCCTCTGTGAGGAGCTGATGCGGAA 712  
 QY 972 TAAACAAGGTATTTCTGTATCAGAAATTTCTCACCGCAGGAGACTGGCAAGTATCAT 1031  
 Db 713 CAATGATCCGATCCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 772  
 QY 1032 CAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1091  
 Db 773 TGACTTTGTGAGCAAGATGATGAGAGATATGAAATPACTGGCATACAGCAACTGG 832  
 QY 1092 TGAGATCTCTAACACGACACTTTTGACAGCTCTCAGGAGACGAAATGGGAGATGCAAG 1151  
 Db 833 GGGCATCAACCAACCATTTCTTGTGATGATGATGATGATGATGATGATGATGATGAT 892  
 QY 1152 GCTCAACAGAGCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1211  
 Db 893 TCTTTGCTGCTTAACTTCTCAACGCTTGAGATGTTTCTGTGATCATCAGAACTGCG 952  
 QY 1212 T 1212  
 Db 953 T 953  
 XX  
 RESULT 15  
 ACC62500  
 ID ACC62500 standard; cDNA; 1804 BP.  
 XX  
 AC ACC62500;  
 XX  
 DT 23-UTR-2003 (first entry)  
 DE Human secreted protein #26 coding sequence SEQ ID 36.  
 XX  
 XX Human; cardiac; hypotensive; hypertensive; cytosolic; antiinflammatory;  
 KM antisthmatic; osteoprotective; antihypertensive; antineoplastic;  
 KM fungicide; virucide; antiparasitic; prozoocidal; antianemic;  
 KM antitumor; immunosuppressive; antineuritic; antiparasitic;  
 KM haemostatic; antidiabetic; antiallergic; antigout; antiarrhythmic;  
 KM antianginal; nephrotoxic; neuroleptic; neuroprotective; noctropic;  
 KM antiparkinsonian; tranquilizer; anorectic; antitumor; hepatotoxic;  
 KM dermatological; gene therapy; neoplastic disease; blood disorder;  
 KM cardiovascular disorder; respiratory disorder; cancer;  
 KM endocrine disorder; reproductive system disorder; infection; arthritis;  
 KM asthma; autoimmune disease; immune deficiency; psoriasis;  
 KM diabetes mellitus; allergy; Paget's disease; gout; osteoporosis;  
 KM renal disorder; Alzheimer's disease; Parkinson's disease; schizophrenia;  
 KM attention deficit disorder; obsessive compulsive disorder; wound healing;

AC ACC62479;  
 DT 23-JUN-2003 (first entry)  
 DE Human secreted protein #5 coding sequence SEQ ID 15.  
 XX

KM Human; Cardiant; hypotensive; hypertensive; cytostatic; antiinflammatory;  
 KM antihistaminic; osteopathic; antithyroid; gynecological; antibacterial;  
 KM fungicide; virucide; antiparasitic; protozoacide; antianemic;  
 KM antiallergic; immunosuppressive; antineumatic; antiparasitic;  
 KM haemostatic; antidiabetic; antiallergic; antigout; antirhythmic;  
 KM antianemic; nephrotoxic; neuroleptic; neuroprotective; nootropic;  
 KM antiparkinsonian; tranquilizer; anorectic; antitumor; hepatotropic;  
 KM dermatological; gene therapy; neoplastic disease; blood disorder;  
 KM cardiovascular disorder; respiratory disorder; cancer;  
 KM endocrine disorder; reproductive system disorder; infection; arthritis;  
 KM asthma; autoimmune disease; immune deficiency; psoriasis;  
 KM diabetes mellitus; allergy; Paget's disease; gout; osteoporosis;  
 KM renal disorder; Alzheimer's disease; Parkinson's disease; schizophrenia;  
 KM attention deficit disorder; obsessive compulsive disorder; wound healing;  
 KM gene; ss.

OS Homo sapiens.  
 XX WO20029066-A2.  
 XX 12-DEC-2002.  
 PD 05-JUN-2002; 2002WO-US017699.  
 XX 06-JUN-2001; 2001US-0295869P.  
 PR 11-JUN-2001; 2001US-0304121P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX

PI Ruben SM, Bell A, Birse CE, Komatsoulis G, Choi GH, Olsen H;  
 PI Ni U, Baker KP;  
 XX WPI; 2003-140609/13.  
 DR P-PSDB; ABR54282.  
 XX

PT Novel human secreted polypeptides and polynucleotides for diagnosing and  
 PT treating neural, immune system, muscular, reproductive, gastrointestinal,  
 PT cardiovascular, renal, proliferative disorders and cancerous diseases.  
 XX

PS Claim 1; Page 539-540; 605pp; English.  
 XX

XX The present invention relates to novel human secreted proteins (ABR54278-  
 CC ABR54331) and their coding sequences (ACC62475-ACC62528). The proteins  
 CC and their coding sequences are useful for treating, preventing,  
 CC diagnosing and/or prognosing neoplastic diseases, blood disorders,  
 CC cardiovascular disorders (e.g. cardiomyopathy, hypertension,  
 CC hypotension), respiratory disorders (e.g. lung cancer, pneumonia,  
 CC bronchitis, asthma), endocrine disorders (e.g. Addison's disease,  
 CC Cushing's syndrome, hyperpituitarism, pituitary dwarfism, acromegaly,  
 CC thyroiditis), reproductive system disorders (e.g. premenstrual syndrome,  
 CC polycystic ovary syndrome), infectious diseases caused by bacteria,  
 CC fungal, viral, parasitic, protozoal, and/or blood-related disorders and  
 CC infections, leukopenia, leukaemia, arthritis, asthma, autoimmune  
 CC diseases, rheumatoid arthritis, immune deficiency, psoriasis,  
 CC haemophilia, diabetes mellitus, allergies, and bone cancers, Paget's  
 CC disease, gout, osteoporosis, arrhythmia, angina, prostate cancer, renal  
 CC disorders, urolithiasis, Alzheimer's disease, Parkinson's disease,  
 CC schizophrenia, attention deficit disorder, obsessive compulsive  
 CC pneumonia, obesity, golfer, ulcerative colitis, hepatitis, They are also  
 CC useful for stimulating epithelial cell proliferation and basal  
 CC keratinocytes for wound healing, and to stimulate hair follicle  
 CC production and healing of dermal wounds  
 XX

XX Sequence 1804 BP; 405 A; 474 C; 457 G; 465 T; 0 U; 3 Other;  
 SQ

Query Match 11.1%; Score 162.6; DB 7; Length 1804;  
 Best Local Similarity 52.5%; Pred. No. 1.5e-41;

Matches 410; Conservative 0; Mismatches 359; Indels 12; Gaps 2;

|    |      |   |      |
|----|------|---|------|
| QY | 441  | CCATCAGTGGGCTGTTTTCATGAGAGTCCCGAAGAAACATTAATTAACCTTTCA    | 500  |
| DB | 158  | CCCTCGGCGCTGGGCGCTCTCCACAGAGAGTGCCCGCCCAACACTTCTTGACCA    | 217  |
| QY | 501  | TAAACAGTATTAATCTTTGTTCACTGACCTGACGTTTACAGAGCATTCACCTGGC   | 560  |
| DB | 218  | CGGCGCGGAGTCCGCTCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT    | 277  |
| QY | 561  | ACTACTACCAACCAATCTGAGAGCACTTAAGCCGAAGTCACTTCACTTCACTT     | 620  |
| DB | 278  | GCTTCCTCTCACTGAGTGGCTGCGGAGCCGCTATCTGGCC---CGCGGTCCTCCGC  | 334  |
| QY | 621  | TTTTCAGTCCAAAACAAGCTTAGAAAAAAGCTTCTGCTGCTGTTATGTAACGTACA  | 680  |
| DB | 335  | CATGAAACGCGGAGAGTGGCGCGCGGCTACAGCGCGCTCTTACTGACATCACA     | 394  |
| QY | 681  | CTGAGACCCACCACTGACAGAGAGCACTATGTCGCGAGTGAAGTCAATGAGGT     | 740  |
| DB | 395  | CTGGAGCTGCGAGCGGACCGGACCGCTACGTGCGGAGCTCATGCGCACATCCGCT   | 454  |
| QY | 741  | CGATTCCTATGTAATGTTTACGAAACAAAGACTCCCT-----CAGCAGCTGA      | 791  |
| DB | 455  | AGACTCCTACGGGAAATGCTGAGAAATCGGAGCTGCTTACCGCGGCTACAGAGAC   | 514  |
| QY | 792  | AAATCCAGCTTATGATGATGCGATGCTTTTATAGATCATGACAGTAAATTTAT     | 851  |
| DB | 515  | AGCCAGGCGCACACCGAGAGTCCAGAGCTCTTGGCTTTCTGCTCCGTATTAATGCA  | 574  |
| QY | 852  | CCTAGCTTTTGAAGATGCAATTTGTGATGATACATCACTGAGAAATTTGAGGCCACT | 911  |
| DB | 575  | CTGGCGCTGGAAGAAATGCAATCTGTAAGACATGACAGAAAAAAGTGGCGTCCAT   | 634  |
| QY | 912  | GAAATGAGGAGTGTGCTGCTGATATTAAGGATGCCAGATCAAGACTGCTTCCAG    | 971  |
| DB | 635  | GCACTGAGGAGTGTGCTGCTGATGATGATGATGATGATGATGATGATGATGATG    | 694  |
| QY | 972  | TAAACAAAGTGTATTTCTGATCAGATTTTCTGACCCAGGAGAACTGGCAATTAAT   | 1031 |
| DB | 695  | CAATCACTCCGCTATCTGATGATGATGATGATGATGATGATGATGATGATGATG    | 754  |
| QY | 1032 | CAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG    | 1091 |
| DB | 755  | TGACTTTCTGACCAAAATGATGAGAGTATGAAATCTGCGCATCAACCAACTGG     | 814  |
| QY | 1092 | TGAGATCTCAACCAAGCACTTCTGACAGCTCTCAGGGAACGGAATGGGAGTCCAGA  | 1151 |
| DB | 815  | GCGCATCACCAACCAATTTCTTCTGATGATGATGATGATGATGATGATGATGATG   | 874  |
| QY | 1152 | CGTCAACCAAGCAATTAATGATGATGATGATGATGATGATGATGATGATGATG     | 1211 |
| DB | 875  | TCTTTGCTGCTTACTACTGATGATGATGATGATGATGATGATGATGATGATG      | 934  |
| QY | 1212 | T 1212  |      |
| DB | 935  | T 935   |      |

#### RESULT 14

ACC62499 standard; cDNA; 1812 BP.

ACC62499;

23-JUN-2003 (first entry)

Human secreted protein #25 coding sequence SEQ ID 35.

Human; Cardiant; hypotensive; hypertensive; cytostatic; antiinflammatory;  
 antihistaminic; osteopathic; antithyroid; gynecological; antibacterial;  
 fungicide; virucide; antiparasitic; protozoacide; antianemic;  
 antiallergic; immunosuppressive; antineumatic; antiparasitic;



Db 848 CCATCTGTAAAGACTATGACAGAAAACCTGGGCTCCCATGACACTGGGCGCTGTGC 907  
Qy 929 CTGTATATTACGATATCCCCAGCATCAACAATCTGGCTTCCAACTAAACAAAGTCTATTTC 988  
Db 908 CCGTGTACCGGGGTTCCTCCCTGTGTGAGGACTGGAGCCGGAACAATCACTCCGTCATCC 967  
Qy 989 TTGTATGAGATTTTTCACCCCGAGGAGCTGGCAATTCATCAACAGCTGGATTCG 1048  
Db 968 TGATTGAGATTTTGTAGCTCTCTCTAGAAAGCTGGCAGATTATTGACTTTTGTGACAA 1027  
Qy 1049 ATGACAGATTGTATGAGCCTATGTAGAAATGGAAGGGTGAATCTTAAACAGC 1108  
Db 1028 ATGATGAGAGATATATGAAATACCTGGCATTAACAGCAACTGGGGGCATCACCAACCAAT 1087  
Qy 1109 GACTTTCGACGCTCTAGGGAACGGAATGGGAGAGCGCAAGCTCAACCGAGCAAT 1168  
Db 1088 TTCTTTGTGATAGCTGTGAAGATGGGAGTGGGAGTAATGATCTTTGTGCTTAAT 1147  
Qy 1169 ACATGATGATTTGATGTATGTGTGACCAAGGTGTGGCT 1212  
Db 1148 ACCTCAAGCGCTCGAGTGTCTGTCTGTGACTACGAACTGGCT 1191  
RESURF 11  
ABA05333  
ID ABA05333 standard; cDNA; 2557 BP.  
AC ABA05333;  
XX  
XX 26-FEB-2002 (first entry)  
DT  
XX  
XX Human fucosyltransferase family member 32132 encoding cDNA.  
DE  
XX  
XX Human; fucosyltransferase family member 32132; cytostatic; antitumour;  
KM antithrombotic; hypotensive; antiarrhythmic; vasoregic;  
KM antidiabetic; antiarthritic; immunosuppressive; antithyroid; antitumor;  
KM antiinflammatory; ophthalmological; antiallergic; antiasthmatic;  
KM nephrotoxic; dermatological; antidiabetic; vulnery; haemostatic;  
KM vitruide; antibacterial; noctropic; neuroprotective; antiparkinsonian;  
KM anticonvulsant; analgesic; anorectic; metabolic; immunomodulator;  
KM proliferation; differentiation; cancer; apoptosis; leukaemia; arthritis;  
KM cardiovascular disease; diabetes mellitus; ulcerative colitis;  
KM Crohn's disease; glomerulonephritis; hepatocellular necrosis;  
KM homeostasis; meningitis; multiple sclerosis; pain; gene therapy; ss.  
XX  
XX Homo sapiens.  
OS  
XX Key Location/Qualifiers  
FH CDS 236..1714  
FT /\*tag= a  
FT /product= "fucosyltransferase family member 32132"  
XX  
XX WO200183721-A2.  
FN  
XX 08-NOV-2001.  
PD  
XX 27-APR-2001; 2001WO-US013805.  
PF  
XX 28-APR-2000; 2000US-0200604P.  
PR  
XX (MILL-) MILLENNIUM PHARM INC.  
PA  
XX  
XX Meyers RA, Williamson M;  
PI  
XX WPI, 2002-041492/05.  
DR P-PSDB; AAM47905.  
XX  
XX Novel isolated 32132 nucleic acid encoding 32132 polypeptide, a novel  
PT member of fucosyltransferase family, useful for treating atherosclerosis,  
PT multiple sclerosis, acute glomerulonephritis, Raynaud's disease, pain.  
XX  
XX Claim 1; Page 105-107; 125sp; English.

XX The invention relates to an isolated 32132 nucleic acid (I), encoding a  
CC member of the fucosyltransferase family comprising: (a) nucleotide  
CC sequence having 80% identity to sequence S1 (ABA05333) or S3 (ABA05334);  
CC (b) comprising fragment of S1 or S3; (c) encoding a polypeptide (II)  
CC comprising a fully defined sequence S2 (AAM47905) of 492 amino acids or  
CC its fragment; or (d) encoding naturally occurring allelic variant of S2.  
CC (1) has cytostatic, antitumor, antithrombotic, hypotensive,  
CC antiarrhythmic, vasoregic, antidiabetic, antiarthritic,  
CC immunosuppressive, antithyroid, antitumor, antiinflammatory,  
CC ophthalmological, antiallergic, antiasthmatic, nephrotoxic,  
CC dermatological, antidiabetic, vulnery, haemostatic, vitruide,  
CC anticonvulsant, analgesic, anorectic, metabolic and immunomodulator  
CC activity (I) and (II) are useful for treating proliferative and/or  
CC differentiative disorders of the colon, including adenoma and colorectal  
CC carcinogenesis, the liver, including nodular hyperplasia and adenomas,  
CC breast, including epithelial hyperplasia and sclerosing adenosis, lung,  
CC including bronchogenic carcinoma and neuroendocrine tumour. The  
CC proliferative disorders treated also include haematopoietic neoplastic  
CC disorders such as acute promyeloid leukaemia, acute myelogenous  
CC leukaemia. (I) and (II) are also useful for treating cardiovascular  
CC diseases, disorders of blood vessels, immunological disorders such as  
CC diabetes mellitus, arthritis, ulcerative colitis, Crohn's disease,  
CC asthma, kidney disorders e.g., polycystic kidney disease, acute  
CC glomerulonephritis, colon disorders, disorders of liver such as disorders  
CC associated with accumulation of fibrous tissue, hepatocellular necrosis  
CC or injury induced by agents including processes which disturb  
CC homeostasis, disorders of brain such as intracranial haemorrhage,  
CC bacterial or viral meningitis, neurodegenerative disorder, glioma,  
CC multiple sclerosis, metabolic or pain disorders. (I) can be used in gene  
CC therapy  
XX  
SQ Sequence 2557 BP; 540 A; 690 C; 684 G; 641 T; 0 U; 2 Other;  
Query Match 12.4%; Score 180.8; DB 6; Length 2557;  
Best Local Similarity 51.8%; Pred. No. 2.1e-47;  
Matches 489; Conservative 0; Mismatches 442; Indels 13; Gaps 3;  
Qy 278 GGTGTGTCCTCCGCTGACGAGGGGAGAGCTGGAGGTTAGGCCAATGTGAGACAGATGCTTT 337  
Db 487 GCATTTCCCACTTCCCGGAGAGCTCGAGGCA-TGAGTGTGCGGGCGGCGTGC 545  
Qy 338 TCTTACACATCAACCGAGCTTACTTCATCATCATGACCAAGCAATCTCTTTATG 397  
Db 546 TGGCGTCCCGGAACCCCGAGGCGTGAAGGACTCGCGAGCGCGCGTGTCTTCTACG 605  
Qy 398 GTACTGACTTTAATATAGATAGCTTACTCTGCTGGAAAGCCCATGATGCTGGGCTG 457  
Db 606 GCAAGATTTCCGCGGTGCGCGCGCGCGCTGCGCTGCGGACGACAGACTGGCGC 665  
Qy 458 TTTTATGAAAGTCCCGAAAAACAATTATAGCTCTTTTAAACAGATGATTA 517  
Db 666 TCTCCACGAGAGTGCCTTCAACAATCTTGTGAGCAACGCGCGGCAATCCGCC 725  
Qy 518 TGTTCAACTACACTGCGAGCTTCAGAGGCAATTCGCACTGCACTACCAATCT 577  
Db 726 TCTTCATCTTACTTCACTTCACTGCGCACTGCGATTAACCGCTGTGCTGAGTGC 785  
Qy 578 TGGAGACATTGAAGTCTGAAGTCACTCCGATACCTTATCTTTGCACTCAAAAA 637  
Db 786 TGCCCGGAGCCGGCTATCT--GGCGCGCGGTGCTCCCGCCATGGAAGCGCGAGT 842  
Qy 638 AGCTTGAAGAAAGACTGTCTCGCTGGGTGTATGTACATGAGTGAATGCCACCATAG 697  
Db 843 GGGCGCGCGCGCTAGCGCGCGGTCTTATGTGCACTCACTGCAACGTCGACAGGG 902  
Qy 698 ACAGGAGACTATGTTCCGAGCTGATGACTTATGATGAGAGTGCATCTTATGAT 757  
Db 903 ACCGGAGCGCTAGCGTGGCGGCTCATGCGGCAATCCCGTGAAGCTCTTAAGGAAAT 962  
Qy 758 GTTTACAAACAAAGCTCTCT-----CAGAGCTGAAAAATCAAGCTCTATAG 808

QY 161 ATGACATACAAAAATGAGAGACACCTACGCTATTATTCATTCTTAGAAGAG 220  
DB 245 ATGACATACAAAAATGAGAGACACCTACGCTATTATTCATTCTTAGAAGAG 304  
QY 221 GATTGACCTTCAACAGAAAAAATGGAATTGAGACGCTACCCCTATTATGCTGCT 280  
DB 305 GATTGACCTTCAACAGAAAAAATGGAATTGAGACGCTACCCCTATTATGCTGCT 364  
QY 281 GGTCCCGCTGACGGGGGAGACTGGAGGTTTGCCCATGTGAGACAAATGCTTTCT 340  
DB 365 GGTCCCGCTGACGGGGGAGACTGGAGGTTTGCCCAATGTGAGACAAATGCTTTCT 424  
QY 341 TCACCATCAACGGGACCTTCCTCATCATCATGACCAAGCATTCCTCTTATG 398  
DB 425 TCACCATCAACGGGACCTTCCTCATCATCATGACCAAGCATTCCTCTTATG 482

RESULT 10  
ABA05334  
ID ABA05334 standard; cDNA; 1479 BP.  
AC ABA05334;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
XX

Human fucosyltransferase family member 32132 coding sequence.

Human, fucosyltransferase family member 32132; cytosolic; antitumour;  
KM antiatherosclerotic; hypotensive; antiarrhythmic; vasotropic;  
KM antidiabetic; antiarthritic; immunosuppressive; antihypoid; antilucer;  
KM antiinflammatory; ophthalmological; antiallergic; antiaesthetic;  
KM nephrotoxic; dermatological; antidiarrhetic; vulnery; haemostatic;  
KM viricide; antibacterial; neurotropic; neuroprotective; antiparkinsonian;  
KM anticonvulsant; analgesic; anorectic; metabolic; immunomodulator;  
KM proliferation; differentiation; cancer; apoptosis; leukemia; arthritis;  
KM cardiovascular disease; diabetes mellitus; ulcerative colitis;  
KM Crohn's disease; glomerulonephritis; hepatocellular necrosis;  
KM homeostasis; meningitis; multiple sclerosis; pain; gene therapy; ss.

Homo sapiens.

Location/Qualifiers  
Key 1..1479  
FT /\*tag= a  
FT /product= "fucosyltransferase family member 32132"  
PN

W0200183721-A2.

08-NOV-2001.

27-APR-2001; 2001WO-US013805.

28-APR-2000; 2000US-0200604P.

(MILL-) MILLENNIUM PHARM INC.

Meyers RA, Williamson M;

WPI; 2002-041492/05.

P-PsDB; AAM47905.

Novel isolated 32132 nucleic acid encoding 32132 polypeptide, a novel  
member of fucosyltransferase family, useful for treating atherosclerosis,  
multiple sclerosis, acute glomerulonephritis, Raynaud's disease, pain.

Claim 1; Page 105-106; 125pp; English.

The invention relates to an isolated 32132 nucleic acid (I), encoding a  
member of the fucosyltransferase family comprising: (a) nucleotide  
sequence having 80% identity to sequence S1 (ABA05333) or S3 (ABA05334);  
(b) comprising fragment of S1 or S3; (c) encoding a polypeptide (II)  
comprising a fully defined sequence S2 (AAM47905) of 492 amino acids or  
its fragment; or (d) encoding naturally occurring allelic variant of S2.

(I) has cytosolic, antitumour, antiatherosclerotic, hypotensive,  
CC immunosuppressive, antidiabetic, antidiarrhetic, antiallergic,  
CC immunosuppressive, antihypoid, antilucer, antiinflammatory,  
CC ophthalmological, antidiarrhetic, vulnery, haemostatic, viricide,  
CC dermatological, neurotropic, neuroprotective, antiparkinsonian,  
CC antibacterial, anorectic, metabolic and immunomodulator  
CC anticonvulsant, analgesic, anorectic, metabolic and immunomodulator  
CC activity. (I) and (II) are useful for treating proliferative and/or  
CC differentiative disorders of the colon, including adenoma and colorectal  
CC carcinogenesis, the liver, including nodular hyperplasia and adenomas,  
CC breast, including epithelial hyperplasia and sclerosing adenosis, lung,  
CC including bronchogenic carcinoma and neuroendocrine tumour. The  
CC proliferative disorders treated also include haematopoietic neoplastic  
CC disorders such as acute promyeloid leukemia, acute myelogenous  
CC leukemia. (I) and (II) are also useful for treating cardiovascular  
CC diseases, disorders of blood vessels, immunological disorders such as  
CC diabetes mellitus, arthritis, ulcerative colitis, Crohn's disease,  
CC asthma, kidney disorders e.g., polycystic kidney disease, acute  
CC glomerulonephritis, colon disorders, disorders of liver such as disorders  
CC associated with accumulation of fibrous tissue, hepatocellular necrosis  
CC or injury induced by agents including processes which disturb  
CC homeostasis, disorders of brain such as intracranial haemorrhage,  
CC bacterial or viral meningitis, neurodegenerative disorder, glioma,  
CC multiple sclerosis, metabolic or pain disorders. (I) can be used in gene  
CC therapy  
XX

Sequence 1479 BP; 280 A; 453 C; 459 G; 287 T; 0 U; 0 Other;

Query Match 12.4%; Score 180.8; DB 6; Length 1479;

Best Local Similarity 51.8%; Pred. No. 1.5e-47;

Matches 499; Conservative 0; Mismatches 442; Indels 13; Gaps 3;

QY 278 GGTGTGCCCGGTGACGGGGGAGAGTGGAGCTTATGAGCAATGCTTGT 337  
DB 252 GGTATTCGCCCACTCCCGGAGAGCTCGAGCGCA-TGAGTGTGCCCGGGCGCG 310  
QY 338 TCTTCAACATCAACGGGACCTACCTCATCATCATGACCAAGCATTCCTTTATG 397  
DB 311 TGGCGTCCGGAGCCCGGAGGCTGAGGAGCTCGGAGCCGCGCTGCTTTACG 370  
QY 398 GTACTGACTTTAATAGATAGTACTTCTGCTCGGAAAGCCCATATGATCGGCTG 457  
DB 371 GCACAGACTTCGGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 430  
QY 458 TTTTTCATGAAGAGTCCCGGAAAAAATTTAAGCTCTTTCAATAAACGATTA 517  
DB 431 TCTTCACAGAGAGTGGCGGCTCAACCACTTCTGTGAGCAAGCGCGCGCGCG 490  
QY 518 TGTTCATACACTGCTCCAGCTTCAGAGGAGCTCCACTTGCACATAACCAAT 577  
DB 491 TCTTCAATTTACCTCCACCTTCAGTCCGCACTGAGATACCGCTGCTGAGAG 550  
QY 578 TGGAGGACTTGAAGTCTGAAGTACTTCGATACCTTCTTTGAGTCCAAAAA 637  
DB 551 TGGCGGAGCCCGCTATCT--GCGCGCGCGGAGCTCCGCCATGGAACGCGGAG 607  
QY 638 AGCTTAGAAAAAGATTGCTGCTGGGTGTATGACAGCTGAGACCCATCAG 697  
DB 608 GGGCGCGCGCGCTACAGCGCGCGCTCTATCTGAGTACACATGAGAGCGG 667  
QY 698 ACAGGAGAGCTATGTTGCGAGCTGATGACTTACATGAGGTGATTTCTATG 757  
DB 668 ACCGGAGCGCTACGTCGCGGAGCTCATGCCCATCCGCTGAGCTCTTACGG 727  
QY 758 GTTTAGAAACAAAGCTCCCT-----CAGAGCTGAAAATCAGCTCTTATG 808  
DB 728 GCCTGAGATGGAAGCTGCTTACCGCGGAGCTTACAGAACAGCCAGCCAC 787  
QY 809 ATGCCATGCTTTTATAGATCATTGACAGATATGATTTATCTTATGAGATG 868  
DB 788 AGGATCAAGCTCTTGGCTTTCTTCCCGCTATTAATGTCACATTTGGCCCTG 847  
QY 869 CAGTTGTATGACTACACTGAGAGGTTTGAAGGCCACTGAATCTGGGGGTATG 928



PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244517P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
XX  
PA (ROSE/) ROSEN C A.  
PA (RUBE/) RUBEN S M.  
XX (BARA/) BARASH S C.  
PI Rosen CA, Ruben SM, Barash SC;  
XX  
XX WPI: 2003-128199/12.  
DR P-PSDB; ABU12557.  
XX  
XX Isolated nucleic acid molecules encoding musculoskeletal system  
PT associated polypeptides, useful for detecting disorders, e.g. cancer.  
XX  
PS Claim 1; SEQ ID NO 187; 321pp; English.  
XX  
XX The invention describes an isolated nucleic acid molecule comprising a  
CC sequence encoding musculoskeletal system associated polypeptides useful  
CC for detecting disorders, e.g., cancer or cancer metastases, in animals or  
CC humans. The nucleic acid: stimulates re-vascularisation of ischaemic  
CC tissues associated with conditions such as thrombosis, arteriosclerosis,  
CC and other cardiovascular conditions; treats wounds due to injuries,  
CC burns, post-operative tissue repair, and ulcers; stimulates angiogenesis  
CC and limb regeneration; stimulates neuronal growth; can treat and prevent  
CC neuronal damage occurring in certain disorders or neurodegenerative  
CC conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-  
CC related complex; stimulates chondrocyte growth, thus they can be used to  
CC enhance bone and periodontal regeneration and aid in tissue transports or  
CC bone grafts; prevents skin aging due to sunburn by stimulating  
CC keratinocyte growth; prevents hair loss, since RGF family members  
CC activate hair-forming cells and promotes melanocyte growth; stimulates  
CC growth and differentiation of hematopoietic cells and bone marrow cells  
CC when used in combination with other cytokines; maintains organs before  
CC transplantation or for supporting cell culture of primary tissues;  
CC induces tissue of mesodermal origin to differentiate in early embryos;  
CC increases or decreases the differentiation or proliferation of embryonic  
CC stem cells, besides, hematopoietic lineage; modulates mammalian  
CC characteristics, such as, body height, weight, hair colour, eye colour,  
CC skin, percentage of adipose tissue, pigmentation, size, and shape (e.g.,  
CC cosmetic surgery); modulates mammalian metabolism; changes mammal's metal  
CC state or physical state by influencing biorhythms, cardiac rhythms,  
CC depression, tendency for violence, tolerance for pain, reproductive  
CC capabilities, hormonal or endocrine levels, appetite, libido, memory, or  
CC stress; increases or decreases storage capabilities, fat content, lipid,  
CC protein, carbohydrate, vitamins, minerals, cofactors or other nutritional  
CC components. This sequence encodes a novel human musculoskeletal system  
CC antigen. Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?DocID=20020147140  
XX  
SQ Sequence 378 BP; 107 A; 90 C; 84 G; 96 T; 0 U; 1 Other;  
Query Match 22.2%; Score 322.8; DB 7; Length 378;  
Best Local Similarity 96.5%; Pred. No. 1.4e-94;  
Matches 363; Conservative 0; Mismatches 8; Indels 5; Gaps 3;

Db 61 GTCCTGAAGTCACTCCGATACCTAGTCTCTTGAGTCCAAAACAAAGCTTAGAAAAGA 120  
QY 652 CTGTGCTCCGCTGGTGTATGTATGATCAGTCAAGTGTGACCCACCATAGACAGGACAGCTAT 711  
Db 121 CTGTGCTCCGCTGGTGTATGTATGATCAGTCAAGTGTGACCCACCATAGACAGGACAGCTAT 180  
QY 712 GTTTCGGAGCTGATATGATATGATGAGGTGCATTCCTATGGTGAATGTTTACGAAACAA 771  
Db 181 GTTTCGGAGCTGATATGATATGATGAGGTGCATTCCTATGGTGAATGTTTACGAAACAA 240  
QY 772 GACCTCCCTCAGCAGCTGAAAAATCCAGCCTCTATAGATGCGATGGCTTTATAGATC 831  
Db 241 GACCTCCCTCAGCAGCTGAAAAATCCAGCCTCTATAGATGCGATGGCTTTATAGATC 300  
QY 832 ATTGCACAGTATTAATGTTATCTTACCTTTTATAGATGAGTGTGATGACTATGACT 891  
Db 301 A-TGCACAGTATTAATGTTATCTTACCTTTTATAGATGAGTGTGATGACTATGACT 355  
QY 892 GAGAGTCTGTGAGGC 907  
Db 356 GCGAAGTTCTGTGAGGC 371  
RESULT 8  
ID ADC30780  
XX ADC30780 standard; cDNA; 1061 BP.  
XX  
XX ADC30780;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Human novel cDNA sequence, SEQ ID NO:862.  
XX  
XX Human; diagnostic; drug screening; forensics; gene mapping;  
KW biodeversity assessment; Parkinson's disease; Alzheimer's disease;  
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
KW ulcers; osteoporosis; autoimmune disease; cancer;  
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
KW neuroprotective; antiangiogenic; anticoagulant; thrombolytic; vulnery;  
KW anticancer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
KW gene therapy; chromosome 8; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
PN WO2003029271-A2.  
XX  
PD 10-APR-2003.  
XX  
XX 24-SEP-2002; 2002WO-US030474.  
XX  
XX 24-SEP-2001; 2001US-0324631P.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Dmanac RT;  
XX  
DR WPI: 2003-371981/35.  
DR P-PSDB; ADC31751.  
XX  
PT New polynucleotide and polypeptide useful for diagnosing, preventing or  
PT treating conditions such as neurodegenerative diseases, anemias, platelet  
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
PT cancer.  
XX  
PS Claim 1; SEQ ID NO 862; 1185pp; English.  
XX  
XX The invention relates to 971 novel human cDNA sequences (ADC29919-  
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
CC invention also relates to nucleic acid sequences over 99% identical with  
CC the novel human cDNAs. The invention additionally encompasses expression  
CC vectors and host cells comprising a nucleic acid of the invention; the



```
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI, 2001-451937/48.
DR P-PSDB; ABB03263.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
XX disorders related to the musculoskeletal system including musculoskeletal
XX cancers and also for testing and detection e.g. diagnosis.
XX
XX Claim 1; SEQ ID NO 187; 761bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AAL34669-AAL37666) and proteins
XX (ABB03087-ABB04109) associated with the musculoskeletal system useful for
XX preventing, treating or ameliorating medical conditions e.g. by protein
XX or gene therapy. The genes are isolated from a range of human tissues
XX disclosed in the specification. The nucleic acids, proteins, antibodies
XX and (ant)agonists are useful in the diagnosis, treatment and prevention
XX of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the
XX adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
XX lung, or urogenital; (b) immune disorders e.g. Addison's disease,
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX arthritis and ulcerative colitis; (c) cardiovascular disorders such as
XX myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
XX cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
XX bacterial, fungal and parasitic infections. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 378 BP; 107 A; 90 C; 84 G; 96 T; 0 U; 1 Other;
Query Match 22.2%; Score 323.8; DB 4; Length 378;
Best Local Similarity 96.5%; Pred. No. 1,4e-94;
Matches 363; Conservative 0; Mismatches 8; Indels 5; Gaps 3;
QY 532 GCCACGTTCCAGCAGGACATTCCTCCACTTCCCACTACTACTCCCAATACCTTGAGAGCATTTGA 591
DB 1 GCCACGTTCCAGCAGGACATTCCTCCACTTCCCACTACTACTCCCAATACCTTGAGAGCATTTGA 60
QY 592 GTCCGAAGTCACTCCGATCCTAGTCCCTTGAGAGTCCAAAACAAGCTTAGAAAAAGA 651
DB 61 GTCCGAAGTCACTCCGATCCTAGTCCCTTGAGAGTCCAAAACAAGCTTAGAAAAAGA 120
QY 652 CTTCCTCCGCTGGTATGATGACAGTCACTGACCCACCATCAAGGACAGCTAT 711
DB 121 CTTCCTCCGCTGGTATGATGACAGTCACTGACCCACCATCAAGGACAGCTAT 180
QY 712 GTTCGCGAGCTGATGACATCATGAGGTGATTCCTATGCTGAATGTTAGAAACAA 771
DB 181 GTTCGCGAGCTGATGACATCATGAGGTGATTCCTATGCTGAATGTTAGAAACAA 240
QY 772 GACCTCCCTCAGCAGCTGAAAATCCAGCCTATAGATGCCAGCTTTATAGATC 831
DB 241 GACCTCCCTCAGCAGCTGAAAATCCAGCCTATAGATGCCAGCTTTATAGATC 300
QY 832 ATTGACAGCTATAGCTTTATCCTTGGAGAAAGCACTTGTATGACATCATCAT 891
DB 301 ATTGACAGCTATAGCTTTATCCTTGGAGAAAGCACTTGTATGACATCATCAT 355
QY 892 GAGAAGTTCTGGAGGC 907
DB 356 GGAAGTTCTGGAGGC 371
```

```
ABX57833
ID ABX57833 standard; cDNA; 378 BP.
XX
XX AC ABX57833;
XX
XX DT 26-FEB-2003 (first entry)
XX
XX DE cDNA encoding novel human musculoskeletal system antigen #177.
XX
XX Gene; ss; musculoskeletal system antigen; cancer; metastasis;
XX re-vascularisation; thrombosis; arteriosclerosis; mineral content;
XX cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;
XX post-operative tissue repair; limb regeneration; neuronal growth;
XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
XX AIDS-related complex; chondrocyte growth; bone regeneration;
XX periodontal regeneration; tissue transport; bone graft; skin aging;
XX keratinocyte growth; hair loss; melanocyte growth; cell proliferation;
XX cell growth; organ transplant; cell differentiation; body height; weight;
XX hair colour; eye colour; skin; percentage of adipose tissue;
XX pigmentation; cosmetic surgery; metabolism; biorhythm; cardiac rhythm;
XX depression; tendency for violence; pain; reproductive capability;
XX hormone level; endocrine level; appetite; libido; memory; stress;
XX storage capability; fat content; lipid content; protein content;
XX carbohydrate content; vitamin content; cofactor content;
XX nutritional component.
XX
XX OS Homo sapiens.
XX
XX PN US2002147140-A1.
XX
XX PD 10-OCT-2002.
XX
XX PF 17-JAN-2001; 2001US-00764877.
XX
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180652P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 07-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 11-JUL-2000; 2000US-0217496P.
XX PR 14-JUL-2000; 2000US-0218290P.
XX PR 16-JUL-2000; 2000US-0220963P.
XX PR 26-JUL-2000; 2000US-0220964P.
XX PR 14-AUG-2000; 2000US-0224518P.
XX PR 14-AUG-2000; 2000US-0224519P.
XX PR 14-AUG-2000; 2000US-0225267P.
XX PR 14-AUG-2000; 2000US-0225268P.
XX PR 14-AUG-2000; 2000US-0225270P.
XX PR 14-AUG-2000; 2000US-0225447P.
XX PR 14-AUG-2000; 2000US-0225757P.
XX PR 14-AUG-2000; 2000US-0225758P.
XX PR 22-AUG-2000; 2000US-0226688P.
XX PR 30-AUG-2000; 2000US-0228924P.
XX PR 01-SEP-2000; 2000US-0229287P.
XX PR 01-SEP-2000; 2000US-0229343P.
XX PR 01-SEP-2000; 2000US-0229344P.
XX PR 01-SEP-2000; 2000US-0229345P.
XX PR 05-SEP-2000; 2000US-0229509P.
XX PR 05-SEP-2000; 2000US-0229513P.
XX PR 08-SEP-2000; 2000US-0231413P.
XX PR 21-SEP-2000; 2000US-0234223P.
XX PR 21-SEP-2000; 2000US-0234274P.
XX PR 25-SEP-2000; 2000US-0234597P.
XX PR 27-SEP-2000; 2000US-0235834P.
XX PR 29-SEP-2000; 2000US-0236327P.
XX PR 29-SEP-2000; 2000US-0236367P.
XX PR 29-SEP-2000; 2000US-0236368P.
XX PR 29-SEP-2000; 2000US-0236369P.
XX PR 29-SEP-2000; 2000US-0236370P.
XX PR 02-OCT-2000; 2000US-0236802P.
XX PR 02-OCT-2000; 2000US-0237037P.
XX PR 02-OCT-2000; 2000US-0237038P.
```

RESULT 7

Mon Sep 13 09:12:39 2004

us-10-080-960-3.rng

Page 7

|    |   |
|----|---|
| XX | antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  |
| KW | vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;   |
| KW | cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder |
| KW | neurological disease; infection; human; secreted protein;               |
| KM | musculoskeletal system; ss.   |
| XX |   |
| OS | Homo sapiens.   |
| XX |   |
| PN | WO200155367-A1.   |
| XX |   |
| PD | 02-AUG-2001.  |
| XX |   |
| PF | 17-JAN-2001; 2001WO-US001338.   |
| XX |   |
| XX | 31-JAN-2000; 2000US-0179065P.   |
| PR | 04-FEB-2000; 2000US-0180628P.   |
| PR | 24-FEB-2000; 2000US-0184664P.   |
| PR | 02-MAR-2000; 2000US-0186350P.   |
| PR | 16-MAR-2000; 2000US-0189874P.   |
| PR | 17-MAR-2000; 2000US-0190076P.   |
| PR | 18-APR-2000; 2000US-0198123P.   |
| PR | 19-MAY-2000; 2000US-0205513P.   |
| PR | 07-JUN-2000; 2000US-0209467P.   |
| PR | 28-JUN-2000; 2000US-0214886P.   |
| PR | 30-JUL-2000; 2000US-0215135P.   |
| PR | 07-JUL-2000; 2000US-0216647P.   |
| PR | 07-JUL-2000; 2000US-0216880P.   |
| PR | 11-JUL-2000; 2000US-0217487P.   |
| PR | 11-JUL-2000; 2000US-0217496P.   |
| PR | 14-JUL-2000; 2000US-0218290P.   |
| PR | 26-JUL-2000; 2000US-0220963P.   |
| PR | 26-JUL-2000; 2000US-0220964P.   |
| PR | 14-AUG-2000; 2000US-0224518P.   |
| PR | 14-AUG-2000; 2000US-0224519P.   |
| PR | 14-AUG-2000; 2000US-0225213P.   |
| PR | 14-AUG-2000; 2000US-0225214P.   |
| PR | 14-AUG-2000; 2000US-0225266P.   |
| PR | 14-AUG-2000; 2000US-0225267P.   |
| PR | 14-AUG-2000; 2000US-0225268P.   |
| PR | 14-AUG-2000; 2000US-0225270P.   |
| PR | 14-AUG-2000; 2000US-0225447P.   |
| PR | 14-AUG-2000; 2000US-0225757P.   |
| PR | 14-AUG-2000; 2000US-0225758P.   |
| PR | 14-AUG-2000; 2000US-0225759P.   |
| PR | 18-AUG-2000; 2000US-0225279P.   |
| PR | 22-AUG-2000; 2000US-0226681P.   |
| PR | 22-AUG-2000; 2000US-0226688P.   |
| PR | 22-AUG-2000; 2000US-0227182P.   |
| PR | 23-AUG-2000; 2000US-0227009P.   |
| PR | 30-AUG-2000; 2000US-0228924P.   |
| PR | 01-SEP-2000; 2000US-0229287P.   |
| PR | 01-SEP-2000; 2000US-0229343P.   |
| PR | 01-SEP-2000; 2000US-0229344P.   |
| PR | 01-SEP-2000; 2000US-0229345P.   |
| PR | 03-SEP-2000; 2000US-0229509P.   |
| PR | 05-SEP-2000; 2000US-0229513P.   |
| PR | 06-SEP-2000; 2000US-0230437P.   |
| PR | 06-SEP-2000; 2000US-0230438P.   |
| PR | 08-SEP-2000; 2000US-0231242P.   |
| PR | 08-SEP-2000; 2000US-0231243P.   |
| PR | 08-SEP-2000; 2000US-023144P.  |
| PR | 08-SEP-2000; 2000US-0231413P.   |
| PR | 08-SEP-2000; 2000US-0231414P.   |
| PR | 08-SEP-2000; 2000US-0232080P.   |
| PR | 12-SEP-2000; 2000US-0232081P.   |
| PR | 12-SEP-2000; 2000US-0231688P.   |
| PR | 14-SEP-2000; 2000US-0232397P.   |
| PR | 14-SEP-2000; 2000US-0232398P.   |
| PR | 14-SEP-2000; 2000US-0232399P.   |
| PR | 14-SEP-2000; 2000US-0232400P.   |
| PR | 14-SEP-2000; 2000US-0232401P.   |
| PR | 14-SEP-2000; 2000US-0233063P.   |
| PR | 14-SEP-2000; 2000US-0233064P.   |
| PR | 14-SEP-2000; 2000US-0233065P.   |

|    |             |                 |
|----|-------------|-----------------|
| PR | 14-SEP-2000 | 2000US-02330656 |
| PR | 21-SEP-2000 | 2000US-02342228 |
| PR | 21-SEP-2000 | 2000US-02342744 |
| PR | 25-SEP-2000 | 2000US-02343997 |
| PR | 25-SEP-2000 | 2000US-02343996 |
| PR | 26-SEP-2000 | 2000US-02354484 |
| PR | 27-SEP-2000 | 2000US-02358348 |
| PR | 27-SEP-2000 | 2000US-02358368 |
| PR | 29-SEP-2000 | 2000US-02363512 |
| PR | 29-SEP-2000 | 2000US-02363517 |
| PR | 29-SEP-2000 | 2000US-02363669 |
| PR | 29-SEP-2000 | 2000US-02363668 |
| PR | 29-SEP-2000 | 2000US-02363700 |
| PR | 02-OCT-2000 | 2000US-02368002 |
| PR | 02-OCT-2000 | 2000US-02370378 |
| PR | 02-OCT-2000 | 2000US-02370366 |
| PR | 02-OCT-2000 | 2000US-02370339 |
| PR | 02-OCT-2000 | 2000US-02370404 |
| PR | 13-OCT-2000 | 2000US-02393358 |
| PR | 13-OCT-2000 | 2000US-02393370 |
| PR | 20-OCT-2000 | 2000US-02412212 |
| PR | 20-OCT-2000 | 2000US-02417855 |
| PR | 20-OCT-2000 | 2000US-02417866 |
| PR | 20-OCT-2000 | 2000US-02417877 |
| PR | 20-OCT-2000 | 2000US-02418088 |
| PR | 20-OCT-2000 | 2000US-02418099 |
| PR | 01-NOV-2000 | 2000US-02441826 |
| PR | 01-NOV-2000 | 2000US-02446178 |
| PR | 08-NOV-2000 | 2000US-02464523 |
| PR | 08-NOV-2000 | 2000US-02464525 |
| PR | 08-NOV-2000 | 2000US-02464526 |
| PR | 08-NOV-2000 | 2000US-02464527 |
| PR | 08-NOV-2000 | 2000US-02464528 |
| PR | 08-NOV-2000 | 2000US-02464533 |
| PR | 08-NOV-2000 | 2000US-02464609 |
| PR | 08-NOV-2000 | 2000US-02464610 |
| PR | 08-NOV-2000 | 2000US-02464611 |
| PR | 17-NOV-2000 | 2000US-02492017 |
| PR | 17-NOV-2000 | 2000US-02492080 |
| PR | 17-NOV-2000 | 2000US-02492098 |
| PR | 17-NOV-2000 | 2000US-02492100 |
| PR | 17-NOV-2000 | 2000US-02492116 |
| PR | 17-NOV-2000 | 2000US-02492117 |
| PR | 17-NOV-2000 | 2000US-02492118 |
| PR | 17-NOV-2000 | 2000US-02492144 |
| PR | 17-NOV-2000 | 2000US-02492454 |
| PR | 17-NOV-2000 | 2000US-02492644 |
| PR | 17-NOV-2000 | 2000US-02492656 |
| PR | 17-NOV-2000 | 2000US-02492657 |
| PR | 17-NOV-2000 | 2000US-02492658 |
| PR | 17-NOV-2000 | 2000US-02492659 |
| PR | 17-NOV-2000 | 2000US-02492660 |
| PR | 17-NOV-2000 | 2000US-02492661 |
| PR | 17-NOV-2000 | 2000US-02492662 |
| PR | 17-NOV-2000 | 2000US-02492663 |
| PR | 17-NOV-2000 | 2000US-02492664 |
| PR | 17-NOV-2000 | 2000US-02492665 |
| PR | 17-NOV-2000 | 2000US-02492666 |
| PR | 17-NOV-2000 | 2000US-02492667 |
| PR | 17-NOV-2000 | 2000US-02492668 |
| PR | 17-NOV-2000 | 2000US-02492669 |
| PR | 17-NOV-2000 | 2000US-02492670 |
| PR | 17-NOV-2000 | 2000US-02492671 |
| PR | 17-NOV-2000 | 2000US-02492672 |
| PR | 17-NOV-2000 | 2000US-02492673 |
| PR | 17-NOV-2000 | 2000US-02492674 |
| PR | 17-NOV-2000 | 2000US-02492675 |
| PR | 17-NOV-2000 | 2000US-02492676 |
| PR | 17-NOV-2000 | 2000US-02492677 |
| PR | 17-NOV-2000 | 2000US-02492678 |
| PR | 17-NOV-2000 | 2000US-02492679 |
| PR | 17-NOV-2000 | 2000US-02492680 |
| PR | 17-NOV-2000 | 2000US-02492681 |
| PR | 17-NOV-2000 | 2000US-02492682 |
| PR | 17-NOV-2000 | 2000US-02492683 |
| PR | 17-NOV-2000 | 2000US-02492684 |
| PR | 17-NOV-2000 | 2000US-02492685 |
| PR | 17-NOV-2000 | 2000US-02492686 |
| PR | 17-NOV-2000 | 2000US-02492687 |
| PR | 17-NOV-2000 | 2000US-02492688 |
| PR | 17-NOV-2000 | 2000US-02492689 |
| PR | 17-NOV-2000 | 2000US-02492690 |
| PR | 17-NOV-2000 | 2000US-02492691 |
| PR | 17-NOV-2000 | 2000US-02492692 |
| PR | 17-NOV-2000 | 2000US-02492693 |
| PR | 17-NOV-2000 | 2000US-02492694 |
| PR | 17-NOV-2000 | 2000US-02492695 |
| PR | 17-NOV-2000 | 2000US-02492696 |
| PR | 17-NOV-2000 | 2000US-02492697 |
| PR | 17-NOV-2000 | 2000US-02492698 |
| PR | 17-NOV-2000 | 2000US-02492699 |
| PR | 17-NOV-2000 | 2000US-02492700 |
| PR | 17-NOV-2000 | 2000US-02492701 |
| PR | 17-NOV-2000 | 200             |

| QY       | 762  | ACGAACCAAGACCTCCCTCAGACGCTAAAAATCCAGCCCTCTATGATGCCATGCGCTT    | 821  |
|----------|--|---|------|
| Db       | 300  | ACGAACCAAGACCTCCCTCAGACGCTAAAAATCCAGCCCTCTATGATGCCATGCGCTT    | 359  |
| QY       | 822  | TTATAGATCATTTGCACAGTATTAAGTTTATCCTAGCTTTTGAGATGCAGTTTGATGA    | 881  |
| Db       | 360  | TTATAGATCATTTGCACAGTATTAAGTTTATCCTAGCTTTTGAGATGCAGTTTGATGA    | 419  |
| QY       | 882  | CTTCATCATCTGAGAAATTTCTGAGAGCCCTGAAACCTGGGGGTATCCTCTGATATTACGG | 941  |
| Db       | 420  | CTTCATCATCTGAGAAATTTCTGAGAGCCCTGAAACCTGGGGGTATCCTCTGATATTACGG | 479  |
| QY       | 942  | ATCCCCAGATCATACAGACTGGCTTCCAAAGTAACAAAAGTCTATTCTTGATCAGATT    | 1000 |
| Db       | 480  | ATCCCCAGATCATACAGACTGGCTTCCAAAGTAACAAAAGTCTATTCTTGATCAGATT    | 539  |
| QY       | 1002   | TTCTCACCCTGAGGAGACTGGCAAGTTATCATCATGACAGATGATTTCTGATGACAGATTG | 1059 |
| Db       | 540  | TTCTCACCCTGAGGAGACTGGCAAGTTATCATCATGACAGATGATTTCTGATGACAGATTG | 597  |
| RESULT 5 |  |   |      |
| AC       | ADCC32194/c  |   |      |
| XX       | ADCC32194  | standard; cDNA; 447 bp.                                       |      |
| XX       | 18-DEC-2003  | (first entry)   |      |
| DE       | Human novel cDNA contig sequence, SEQ ID NO:2276.                        |   |      |
| XX       |  |   |      |
| KM       | Human; diagnostic; drug screening; forensics; gene mapping;              |   |      |
| KM       | biodiversity assessment; Parkinson's disease; Alzheimer's disease;       |   |      |
| KM       | neurodegenerative diseases; anemia; platelet disorder; wound; burns;     |   |      |
| KM       | ulcers; osteoporosis; autoimmune disease; cancer;                        |   |      |
| KM       | molecular weight marker; food supplement; antiparkinsonian; nootropic;   |   |      |
| KM       | neuroprotective; anaemic; anticoagulant; thrombolytic; vulnery;          |   |      |
| KM       | antidicer; osteopathic; immunosuppressive; antiinflammatory; cytostatic; |   |      |
| KM       | gene therapy; chromosome 8; ss.  |   |      |
| XX       |  |   |      |
| OS       | Homo sapiens.  |   |      |
| XX       |  |   |      |
| XX       | W02003029271-A2.   |   |      |
| PN       |  |   |      |
| XX       | 10-APR-2003.   |   |      |
| PD       |  |   |      |
| XX       |  |   |      |
| PF       | 24-SEP-2002; 2002MO-US030474.  |   |      |
| XX       |  |   |      |
| XX       | 24-SEP-2001; 2001US-0324631P.  |   |      |
| PR       |  |   |      |
| XX       | (HYSEQ-) HYSEQ INC.  |   |      |
| PA       |  |   |      |
| XX       |  |   |      |
| PI       | Tang TY, Zhang J, Ren F, Xue AJ, Zhao QJ, Wang J, Wehrman T;             |   |      |
| PI       | Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;                 |   |      |
| PI       | Haley-Vicence D, Drmanac RJ;   |   |      |
| XX       |  |   |      |
| XX       | WPI; 2003-371981/35.   |   |      |
| DR       | P-PSDB; ADCC32961.   |   |      |
| XX       |  |   |      |
| PT       | New polynucleotide and polypeptide useful for diagnosing, preventing or  |   |      |
| PT       | treating conditions such as neurodegenerative diseases, anemia, platelet |   |      |
| PT       | disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or   |   |      |
| PT       | cancer.  |   |      |
| XX       |  |   |      |
| PS       | Example 2; SEQ ID NO 2276; 1185bp; English.                              |   |      |
| XX       |  |   |      |
| CC       | The invention relates to 971 novel human cDNA sequences (ADCC29919-      |   |      |
| CC       | ADCC30889) and the polypeptides they encode (ADCC30890-ADCC31860). The   |   |      |
| CC       | invention also relates to nucleic acid sequences over 99% identical with |   |      |
| CC       | the novel human cDNAs. The invention additionally encompasses expression |   |      |
| CC       | vectors and host cells comprising a nucleic acid of the invention; the   |   |      |
| CC       | recombinant production of a polypeptide of the invention; an antibody    |   |      |
| CC       |  |   |      |

CC against polypeptide of the invention; a method of detecting  
CC polynucleotides or polypeptides of the invention; and methods of  
CC identifying a compound which binds to a polypeptide of the invention. The  
CC invention further discloses methods of preventing, treating or  
CC ameliorating a medical condition; kits comprising polynucleotide probes  
CC and/or monoclonal antibodies for carrying out the methods of the  
CC invention; methods for the identification of compounds that modulate the  
CC expression or activity of the polynucleotide and/or polypeptide; and 767  
CC coding sequences corresponding to the cDNA sequences of the invention  
CC (ADCC1861-ADCC2627) and the polypeptides encoded by the contigs (ADCC2628  
CC -ADCC3394). The nucleic acids and polypeptides of the invention are  
CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
CC identification of mutations responsible for genetic disorders or other  
CC traits, for assessing biodiversity, and in producing many other types of  
CC data and products dependent on DNA and amino acid sequences. They are  
CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
CC disease and other neurodegenerative diseases, anaemia, platelet  
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
CC cancer. The nucleic acids may also be used as hybridisation probes or  
CC primers, and in the recombinant production of a protein. The polypeptides  
CC are also useful in generating antibodies, as molecular weight markers,  
CC and as food supplements. The present sequence represents a human contig  
CC sequence used in an example of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences.

| Seq                   | Sequence  | 447 BP;  | 113 A;        | 103 C;      | 101 G; | 130 T; | 0 U; | 0 Other; |
|-----------------------|---|--|---------------|-------------|--------|--------|------|----------|
| Query Match           | 28.5%;  | Score 416.4;   | DB 9;         | Length 447; |        |        |      |          |
| Best Local Similarity | 99.8%;  | Pred. No. 8.8e-125;  |               |             |        |        |      |          |
| Matches 417;          | Conservative  | 0;   | Mismatches 1; | Indels      | 0;     | Gaps   | 0;   |          |
| OY                    | 632   | AAACAAGCTTAGAAAAAGACTTGCCTCCGCTGATGTATGACATGCAGCTGTGACCCAC     | 639           |             |        |        |      |          |
| Db                    | 447   | AAACAAGCTTAGAAAAAGACTTGCCTCCGCTGGGTATGTACATGCAGCTGTATCCAC      | 368           |             |        |        |      |          |
| OY                    | 692   | CATCAACAAGGAGACAGCTTATGTTCCGAGCTTGATGACTTACATGAGGTGATTCCTATG   | 721           |             |        |        |      |          |
| Db                    | 387   | CATCAACAAGGAGACAGCTTATGTTCCGAGCTGTATGACTTACATGAGGTGATTCCTATG   | 328           |             |        |        |      |          |
| OY                    | 752   | GTGAATGTTTACGAAACAAGACCTCCCTCAGCAGCTGAAATAATCCAGCCTATGATG      | 811           |             |        |        |      |          |
| Db                    | 327   | GTGAATGTTTACGAAACAAGACCTCCCTCAGCAGCTGAAATAATCCAGCCTATGATG      | 268           |             |        |        |      |          |
| OY                    | 812   | CCGATGGCTTTTATAGATCATTTGACACAGTATAGTTATTCCTAGCTTTGAGAAATCCAG   | 871           |             |        |        |      |          |
| Db                    | 267   | CCGATGGCTTTTATAGATCATTTGACACAGTATAGTTATTCCTAGCTTTGAGAAATCCAG   | 208           |             |        |        |      |          |
| OY                    | 872   | TTTGATGATGACTACATCATCTGAGAACTTGTGGAGGCCACTGAAACTGGGAGTACTCCCTG | 931           |             |        |        |      |          |
| Db                    | 207   | TTTGATGATGACTACATCATCTGAGAACTTGTGGAGGCCACTGAAACTGGGAGTACTCCCTG | 148           |             |        |        |      |          |
| OY                    | 932   | TATATTACGATATCCCCAGCATCAACAGCTGGCTTCCAAATTAACAAAGTCTATTCCTG    | 992           |             |        |        |      |          |
| Db                    | 147   | TATATTACGATATCCCCAGCATCAACAGCTGGCTTCCAAATTAACAAAGTCTATTCCTG    | 88            |             |        |        |      |          |
| OY                    | 992   | TATCAGAAATTTTCTCAACCCAGGGAAGTGGCAAGTTACATCAGACGACTGGATTCTGA    | 1049          |             |        |        |      |          |
| Db                    | 87  | TATCAGAAATTTTCTCAACCCAGGGAAGTGGCAAGTTACATCAGACGACTGGATTCTGA    | 30            |             |        |        |      |          |
| RESULT 6              |   |  |               |             |        |        |      |          |
| AA134845              |   |  |               |             |        |        |      |          |
| ID                    | AA134845  | standard;  | cDNA;         | 378 BP.     |        |        |      |          |
| XX                    | AA134845;   |  |               |             |        |        |      |          |
| AC                    |   |  |               |             |        |        |      |          |
| DT                    | 08-JAN-2002   | (first entry)  |               |             |        |        |      |          |
| XX                    |   |  |               |             |        |        |      |          |
| XX                    | Human musculoskeletal system related polynucleotide SEQ ID NO 187.    |  |               |             |        |        |      |          |
| XX                    |   |  |               |             |        |        |      |          |
| XX                    | Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral; |  |               |             |        |        |      |          |

```

QY 937 TAGGATCCCCAGCATCAGACGTGGCTTCCAGTAACAAAAGTCTATTCTTGTATCA 996
Db 1272 TAGGATCCCCAGCATCAGACGTGGCTTCCAGTAACAAAAGTCTATTCTTGTATCA 1331
QY 997 GAATTTCTCAGCCCGAGGAAGTGGCAAGTTACATCAGACGACTGGATTCTGTATCAGA 1056
Db 1332 GAATTTCTCAGCCCGAGGAAGTGGCAAGTTACATCAGACGACTGGATTCTGTATCAGA 1391
QY 1057 TTGTATGAGGCTTATGTAGTAAGTGAAGGCTGAAGGCTGAGATCTTAAACGAGCACTTCTG 1116
Db 1392 TTGTATGAGGCTTATGTAGTAAGTGAAGGCTGAGATCTTAAACGAGCACTTCTG 1451
QY 1117 ACAGCTCTCAGGGAAGGGAATGGGAGTGCMAAGCTCAACAGAGCAATTAATCATCAT 1176
Db 1452 ACAGCTCTCAGGGAAGGGAATGGGAGTGCMAAGCTCAACAGAGCAATTAATCATCAT 1511
QY 1177 GCATTTGAGTGTATGTGTGCACCAAGTGTGGGCTAATATCAGGCTTCAGGAAAAGGCG 1236
Db 1512 GCATTTGAGTGTATGTGTGCACCAAGTGTGGGCTAATATCAGGCTTCAGGAAAAGGCG 1571
QY 1237 TTACACCCCAAAAGATGGGAGGCAAGATACCCACTGAGTTGCCAGAGCCCAAGTG 1296
Db 1572 TTACACCCCAAAAGATGGGAGGCAAGATACCCACTGAGTTGCCAGAGCCCAAGTG 1631
QY 1297 TTGGCTTTCCACCACTCCGGAAGTCCACCTTGAGCTTTTGCAGAGATGTGATTTCC 1356
Db 1632 TTGGCTTTCCACCACTCCGGAAGTCCACCTTGAGCTTTTGCAGAGATGTGATTTCC 1691
QY 1357 AGCTTTGAACATTCACAAAGAAAGGCCGACGACTAAGTGGCTGTGTAGAGATCAA 1416
Db 1692 AGCTTTGAACATTCACAAAGAAAGGCCGACGACTAAGTGGCTGTGTAGAGATCAA 1751
QY 1417 AACTTTTCATCTCAAGAGTTTGGGGGCTAGATTCAAGAGACTGA 1461
Db 1752 AACTTTTCATCTCAAGAGTTTGGGGGCTAGATTCAAGAGACTGA 1796

RESULT 4
ABN78606
ID ABN78606 standard; cDNA; 597 BP.
AC ABN78606;
XX
DT 08-JUL-2002 (first entry)
XX
DE Human ORF3553 cDNA, SEQ ID NO:7105.
XX
KV Human; ORF; open reading frame; ORF; drug screening; diagnosis;
KV disease monitoring; cytokine; cell proliferation; cell differentiation;
KV immune modulation; haematopoiesis regulation; tissue growth;
KV angiogenesis; activin; inhibin; chemotactic; chemokine; haemostatic;
KV thrombolytic; tumour inhibition; bodily characteristics; fertility;
KV behaviour; cancer; proliferative disorder; neurological disorder;
KV cardiovascular disease; immune system disorder; organ transplantation;
KV tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KV hypothyroidism; cholesterol ester storage disease; infection; vulnery;
KV vasoprotective; antiproliferative; antidiabetic; cyostatic; nocrotic;
KV neuroprotective; antithrombotic; anticoagulant; thrombolytic;
KV cardiac; hypotensive; anticholesteric; antiinflammatory; immunomodulator;
KV dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
XX
OS Homo sapiens.
XX
PN W0200190366-A2.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001MO-US017076.
XX
PR 24-MAY-2000; 2000US-0206690P.
XX
PA (CURA-) CURAGEN CORP.

```

---

```

XX Leach MD, Shinkens RA;
XX WPI; 2002-106200/14.
DR P-PSDB; ABP34580.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation.
PS Claim 1; Page 2022; 2508BP; English.
XX
CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN78054-
CC ABN78587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antiinfective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases
XX
SQ Sequence 597 BP; 171 A; 145 C; 124 G; 157 T; 0 U; 0 Other;
Query Match 39.8%; Score 581.2; DB 6; Length 597;
Best Local Similarity 99.3%; Pred. No. 1.7e-178;
Matches 594; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 462 TCATGAAGAGTCCCGAAGAAACATTAAGCTTTCTAATAACAGAGTATCCCTGTT 521
Db 1 TCATGAAGAGTCCCG-AAAACATTAAGCTTTCTAATAACAGAGTATCACTTGTT 59
QY 522 CAACTACACTGCGACAGTTCAGACAGCAATCCCACTTCCCACTAATCAATCACTTGA 581
Db 60 CAATACACTGCGACAGTTCAGACAGCAATCCCACTTCCCACTAATCAATCACTTGA 119
QY 582 GAGCATTAAGTCCCGAAGTCACTCCGATTAAGTCTTTCGAGTCCCAAAACAGCT 641
Db 120 GAGCATTAAGTCCCGAAGTCACTCCGATTAAGTCTTTCGAGTCCCAAAACAGCT 179
QY 642 TAGAAAAAGACTTGCTCCGCTGCTGTATGTACAGTCAAGTCAAGCCCAATCAGACAG 701
Db 180 TAGAAAAAGACTTGCTCCGCTGCTGTATGTACAGTCAAGTCAAGCCCAATCAGACAG 239
QY 702 GGCAGCTATGTTGGCGAGCTGATGACTTACATGAGGTGATTCCTATGTTGAATGTTT 761
Db 240 GGCAGCTATGTTGGCGAGCTGATGACTTACATGAGGTGATTCCTATGTTGAATGTTT 299

```

DB 1818 GCCGAGCAGCTAAGTGCTGTGATGAGATCAAACTTTTCATCTCAAGATTG 1877  
QY 1441 GGCTAGATTCAAGACTGA 1461  
DB 1878 GGCTAGATTCAAGACTGA 1898

RESULT 3  
ADC30307  
ID ADC30307 standard; CDNA; 2304 BP.  
XX  
XX ADC30307;  
DT 18-DEC-2003 (first entry)  
XX  
XX Human novel CDNA sequence, SEQ ID NO:389.  
DE  
XX  
XX Human; diagnostic; drug screening; forensics; gene mapping;  
KM biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
KM neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
KM ulcers; osteoporosis; autoimmune disease; cancer;  
KM molecular weight marker; food supplement; antiparkinsonian; nootropic;  
KM neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;  
KM anticancer; osteoporotic; immunosuppressive; antiinflammatory; cytostatic;  
KM gene therapy; chromosome 8; gene, ss.

XX  
XX Homo sapiens.  
OS  
XX  
XX MO2003029271-A2.  
PN  
XX  
XX 10-APR-2003.  
PD  
XX  
XX 24-SEP-2002; 2002MO-US030474.  
PF  
XX  
XX 24-SEP-2001; 2001US-0324631P.  
PR  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Weinman T;  
PI Zhou F, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Dmanac RT;  
PI  
XX  
XX MPI; 2003-371981/35.  
DR  
XX  
XX P-PSDB; ADC31278.

PT New polynucleotide and polypeptide useful for diagnosing, preventing or  
PT treating conditions such as neurodegenerative diseases, anemias, platelet  
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
PT cancer.  
PS  
XX  
XX Claim 1; SEQ ID NO 389; 1185bp; English.

CC The invention relates to 971 novel human CDNA sequences (ADC39919-  
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
CC invention also relates to nucleic acid sequences over 99% identical with  
CC the novel human cDNAs. The invention additionally encompasses expression  
CC vectors and host cells comprising a nucleic acid of the invention; the  
CC recombinant production of a polypeptide of the invention; an antibody  
CC against a polypeptide of the invention; a method of detecting  
CC polynucleotides or polypeptides of the invention; and methods of  
CC identifying a compound which binds to a polypeptide of the invention. The  
CC invention further discloses methods of preventing, treating or  
CC ameliorating a medical condition, kits comprising polynucleotide probes  
CC and/or monoclonal antibodies for carrying out the methods of the  
CC invention; methods for the identification of compounds that modulate the  
CC expression or activity of the polynucleotide and/or polypeptide; and 767  
CC config sequences corresponding to the CDNA sequences of the invention  
CC (ADC31861-ADC32627) and the polypeptides encoded by the configs (ADC32628  
CC -ADC33394). The nucleic acids and polypeptides of the invention are  
CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
CC identification of mutations responsible for genetic disorders or other  
CC traits, for assessing biodiversity, and in producing many other types of  
CC data and products dependent on DNA and amino acid sequences. They are

CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
CC disease and other neurodegenerative diseases, anaemia, platelet  
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
CC cancer. The nucleic acids may also be used as hybridisation probes or  
CC primers, and in the recombinant production of a protein. The polypeptides  
CC are also useful in generating antibodies, as molecular weight markers,  
CC and as food supplements. The present sequence represents a specifically  
CC claimed human CDNA sequence of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcc\_sequences.  
XX  
XX  
SQ Sequence 2304 BP; 639 A; 521 C; 535 G; 609 T; 0 U; 0 Other;  
Query Match 93.0%; Score 1358.6; DB 9; Length 2304;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1361; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 97 AAGAGGTCATGGTTAGCTGGGAAATTGTAAGAGAGACTTTAAAGTTCAGTTG 156  
DB AAGAAATGTCATGGTTAGCTGGGAAATTGTAAGAGAGACTTTAAAGTTCAGTTG 491  
QY 157 CAAGATGACATCAAAATGAGAGACCACTACGCATCTTAATTCATTCTTAAGAA 216  
DB CAAAGATGACATCAAAATGAGAGACCACTACGCATCTTAATTCATTCTTAAGAA 551  
QY 217 GAAGATTGACCTTCAACAGAAAAGAAAATGGGAATTGGACAGCTACCCATTATGCTC 276  
DB GAAGATTGACCTTCAACAGAAAAGAAAATGGGAATTGGACAGCTACCCATTATGCTC 611  
QY 277 TGGTGGTCCCCGCTGACGGGGAGACTGGAGGTTAAGCCAAATGGAGCAATGCTTGT 336  
DB TGGTGGTCCCCGCTGACGGGGAGACTGGAGGTTAAGCCAAATGGAGCAATGCTTGT 671  
QY 337 TTCTTCACCATCAACCGGACCTACCTCATCATGACATGACCAAGACATTCCTCTCTAT 396  
DB TTCTTCACCATCAACCGGACCTACCTCATCATGACATGACCAAGACATTCCTCTCTAT 731  
QY 397 GGTACTGACTTTAACTAATAGTACTTACTCTGCTCGGAAAGCCCATCATCTGGGCT 456  
DB GGTACTGACTTTAACTAATAGTACTTACTCTGCTCGGAAAGCCCATCATCTGGGCT 791  
QY 457 GTTTTCATGAAGAGTCCCCGAAACAAATTAAAGCTCTTTCATTAACAGATGATTAC 516  
DB GTTTTCATGAAGAGTCCCCGAAACAAATTAAAGCTCTTTCATTAACAGATGATTAC 851  
QY 792 GTTTTCATGAAGAGTCCCCGAAACAAATTAAAGCTCTTTCATTAACAGATGATTAC 851  
DB GTTTTCATGAAGAGTCCCCGAAACAAATTAAAGCTCTTTCATTAACAGATGATTAC 911  
QY 517 TTGTTCACTACACTGCGCATGTTACAGAGGCAATTCCTGCACTAATCCCAATAC 576  
DB TTGTTCACTACACTGCGCATGTTACAGAGGCAATTCCTGCACTAATCCCAATAC 911  
QY 577 TTGAGAGCATTGAAGTCTGGAAGTCACTCGATACCTAGTTCCTTGAGGCCAAAAC 636  
DB TTGAGAGCATTGAAGTCTGGAAGTCACTCGATACCTAGTTCCTTGAGGCCAAAAC 971  
QY 637 AAGCTTAGAAAAAGACTTCTCGCTGATGATGATACAGTCACTGAGCCACCATCA 696  
DB AAGCTTAGAAAAAGACTTCTCGCTGATGATGATACAGTCACTGAGCCACCATCA 1031  
QY 697 GAAAGGACAGCTATGTTGCGAGCTGATGATGATGATGATGATGATGATGATGAT 756  
DB GAAAGGACAGCTATGTTGCGAGCTGATGATGATGATGATGATGATGATGATGAT 1091  
QY 1032 GAAGAGGACAGCTATGTTGCGAGCTGATGATGATGATGATGATGATGATGATGAT 1091  
DB GAAGAGGACAGCTATGTTGCGAGCTGATGATGATGATGATGATGATGATGATGAT 1151  
QY 757 TGTTCGAAAAAAGACCTCTCGAGCAGCTGAAATAATCCAGCTCTATGATGATGAT 816  
DB TGTTCGAAAAAAGACCTCTCGAGCAGCTGAAATAATCCAGCTCTATGATGATGAT 1151  
QY 817 GGCTTTTATGATCACTTGCACAGTATAAGTTATTCCTAGCTTTTGAAGATGCAATTGT 876  
DB GGCTTTTATGATCACTTGCACAGTATAAGTTATTCCTAGCTTTTGAAGATGCAATTGT 1211  
QY 877 GATGACTACATCTAGAAAGTTTGGAGGCCACTGAAATCGGGGTATGCTCTATAT 936  
DB GATGACTACATCTAGAAAGTTTGGAGGCCACTGAAATCGGGGTATGCTCTATAT 1271

KM Human: secreted protein; gene therapy; vaccine; treatment; diagnosis;  
 KW GENSET; ss.  
 XX Homo sapiens.  
 OS  
 PN WO200142451-A2.  
 XX  
 XX 14-JUN-2001.  
 PD  
 XX  
 PF 07-DEC-2000; 2000WO-1B001938.  
 XX  
 PR 08-DEC-1999; 99US-0169629P.  
 PR 06-MAR-2000; 2000US-0187470P.  
 XX  
 PA (GEST ) GENSET.  
 PI Dumas Milne Edwards J, Bougueleret L, Jobert S;  
 DR WPI: 2001-367870/38.  
 DR P-PSDB; AAG89164..  
 XX  
 XX Full length GENSET human nucleic acid encoding potentially secreted  
 PT proteins, useful in gene therapy and vaccination against a variety of  
 PT diseases, and for diagnosis of those diseases.  
 PS Claim 7; Page 606-607; 921pp; English.  
 XX  
 CC The invention relates to full length GENSET human nucleic acids encoding  
 CC potentially secreted proteins. The nucleic acids and the polypeptides  
 CC they encode may be used in the prevention, treatment and diagnosis of  
 CC diseases associated with inappropriate GENSET gene expression. For  
 CC example, they be used to treat disorders associated with decreased GENSET  
 CC gene expression by rectifying mutations or deletions in a patient's  
 CC genome that affect the activity of GENSET or by supplementing the  
 CC patient's own production of GENSET polypeptides. Conversely, antisense  
 CC nucleic acid molecules may be administered to down regulate GENSET  
 CC expression by binding with the cells' own genes and preventing their  
 CC expression. The sense and antisense nucleic acids may also be used as DNA  
 CC probes in diagnostic assays to detect and quantitate the presence of  
 CC similar nucleic acid sequences in samples, and hence to determine which  
 CC patients may be in need of restorative therapy. The GENSET polypeptides  
 CC may be used as antigens in the production of antibodies and in assays to  
 CC identify modulators (agonists and antagonists) of GENSET polypeptide  
 CC expression and activity. The present sequence is a GENSET nucleic acid of  
 CC the invention  
 CC  
 XX  
 SQ Sequence 1942 BP; 545 A; 460 C; 462 G; 475 T; 0 U; 0 Other;  
 Query Match 99.0%; Score 1445.8; DB 5; Length 1942;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1458; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db  
 739 ACTGGAGGTTAAGCCAAATGTGAGCAGATGCTTTGTTCTTCAACATCAACCGGACTTAC 798  
 QY  
 361 CTCATCATCATGATGACCAAAAGCATTCCTCTTATGATGATGATCTTAAATATGATAGC 420  
 Db  
 799 CTCATCATCATGATGACCAAAAGCATTCCTCTTATGATGATGATCTTAAATATGATAGC 858  
 QY  
 421 TTACTCTGCTCGGAAAAGCCATCATGATGATGATGATGATGATGATGATGATGATGATG 480  
 Db  
 859 TTACTCTGCTCGGAAAAGCCATCATGATGATGATGATGATGATGATGATGATGATGATG 918  
 QY  
 481 AACATTTAAGCTCTTTCATTAACCAAGTATTAACCTTGTTCATGATGATGATGATGATG 540  
 Db  
 919 AACATTTAAGCTCTTTCATTAACCAAGTATTAACCTTGTTCATGATGATGATGATGATG 978  
 QY  
 541 AGCAGGATTTCCCATCTTCCCATCTAATACCAATACCTTGAAGGATTTGAAGTCCGAA 600  
 Db  
 979 AGCAGGATTTCCCATCTTCCCATCTAATACCAATACCTTGAAGGATTTGAAGTCCGAA 1038  
 QY  
 601 TCATCCGATACCTAGTTCCTTTCAGTCCCAAAAACAGCTTAGAAAAAGACTTGCTCCG 660  
 Db  
 1039 TCATCCGATACCTAGTTCCTTTCAGTCCCAAAAACAGCTTAGAAAAAGACTTGCTCCG 1098  
 QY  
 661 CTGGTGTATGTACAGTCACTGATGACCTGACCCCATCAGACAGGACAGCTTGTCCGAG 720  
 Db  
 1099 CTGGTGTATGTACAGTCACTGATGACCTGACCCCATCAGACAGGACAGCTTGTCCGAG 1158  
 QY  
 721 CTGATGACTTACATCGAGGTCGATTCCTATGATGATGATGATGATGATGATGATGATG 780  
 Db  
 1159 CTGATGACTTACATCGAGGTCGATTCCTATGATGATGATGATGATGATGATGATGATG 1218  
 QY  
 781 CAGCAGCTGAAAAATCCAGCTCTATGATGATGATGATGATGATGATGATGATGATGATG 840  
 Db  
 1219 CAGCAGCTGAAAAATCCAGCTCTATGATGATGATGATGATGATGATGATGATGATGATG 1278  
 QY  
 841 TTAATGTTATTCCTAGCTTTTGAATGACATGATGATGATGATGATGATGATGATGATG 900  
 Db  
 1279 TTAATGTTATTCCTAGCTTTTGAATGACATGATGATGATGATGATGATGATGATGATG 1338  
 QY  
 901 TGAAGGACCTGAAACTGGGGGTAGTCCCTGTATATTAACGATCCCAACATCAGACAG 960  
 Db  
 1339 TGAAGGACCTGAAACTGGGGGTAGTCCCTGTATATTAACGATCCCAACATCAGACAG 1398  
 QY  
 961 TGGCTTCCAGTAACAAAGTGTATTTCTGTATCAGAAATTTTCTCACCCAGGAAATG 1020  
 Db  
 1399 TGGCTTCCAGTAACAAAGTGTATTTCTGTATCAGAAATTTTCTCACCCAGGAAATG 1458  
 QY  
 1021 GCAAGTTACATCAGACGCTGGATTCGATGACAGATTTGATGAGGCTATGTAGAATGG 1080  
 Db  
 1459 GCAAGTTACATCAGACGCTGGATTCGATGACAGATTTGATGAGGCTATGTAGAATGG 1518  
 QY  
 1519 AAGCTGAAGGGT-AGATCTTAACCAAGCACTTCTGACACTCAGAGGAAACGAAATGG 1577  
 Db  
 1141 GGAATGCAAGACGTCACACAGACCAATTAACATGATGATGATGATGATGATGATGATG 1200  
 QY  
 1578 GGAATGCAAGACGTCACACAGACCAATTAACATGATGATGATGATGATGATGATGATG 1637  
 Db  
 1201 AAGGTGTGGGCTAATATCAGGCTTCAGAAAAAGGCTTACACCCAAAGATGGAGGCA 1260  
 QY  
 1639 AAGGTGTGGGCTAATATCAGGCTTCAGAAAAAGGCTTACACCCAAAGATGGAGGCA 1697  
 Db  
 1261 GAAATATCCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320  
 QY  
 1698 GAAATATCCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1757  
 Db  
 1321 CCACCTTTGAGCTCTTTCGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
 QY  
 1758 CCACCTTTGAGCTCTTTCGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 1817  
 Db  
 1381 GCCAGGCACTAAGGTGCTGTTGATAGGATAAAACTTTTCATCTCAAGAGTTTGG 1440



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2004, 16:57:53; Search time 412.627 Seconds

(without alignments)  
15041.697 Million cell updates/sec

Title: US-10-080-960-3

Perfect score: 1461

Sequence: 1 atgaagtcacgcggccctcc.....gcctcattcgaactga 1461

Scoring table: IDENTITY NTC

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: N\_Geneseq\_29Jan04:\*

1: geneseqn1908:\*  
2: geneseqn1908:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length | ID       | Description        |
|------------|--------|-------------|--------|----------|--------------------|
| 1          | 1457.8 | 99.8        | 1669   | ABK53075 | ABK53075 cDNA enco |
| 2          | 1445.8 | 99.0        | 1942   | AAH64767 | AAH64767 Human sec |
| 3          | 1358.6 | 93.0        | 2304   | ADC30307 | ADC30307 Human sec |
| 4          | 581.2  | 39.8        | 597    | ABN78606 | ABN78606 Human ORF |
| 5          | 416.4  | 28.5        | 447    | ADC32194 | ADC32194 Human nov |
| 6          | 323.8  | 22.2        | 378    | AAI34845 | AAI34845 Human mus |
| 7          | 323.8  | 22.2        | 378    | ABX57833 | ABX57833 cDNA enco |
| 8          | 300    | 20.5        | 1061   | ADC30780 | ADC30780 Human nov |
| 9          | 298    | 20.4        | 1550   | ABG67785 | ABG67785 Human rec |
| 10         | 180.8  | 12.4        | 1479   | ABA05334 | ABA05334 Human fuc |
| 11         | 180.8  | 12.4        | 2557   | ABA05333 | ABA05333 Human fuc |
| 12         | 179.8  | 12.3        | 1431   | ABU07657 | ABU07657 Drosophi  |
| 13         | 162.6  | 11.1        | 1804   | ACC62479 | ACC62479 Human sec |
| 14         | 162.6  | 11.1        | 1804   | ACC62499 | ACC62499 Human sec |
| 15         | 162.2  | 11.1        | 1804   | ACC62500 | ACC62500 Human sec |
| 16         | 150.8  | 10.3        | 1328   | ABK11711 | ABK11711 DNA enco  |
| 17         | 139.6  | 9.6         | 676    | AA641344 | AA641344 Human adu |
| 18         | 107.8  | 7.4         | 477    | ACH14839 | ACH14839 Human adu |
| 19         | 107.6  | 7.4         | 4200   | ABU07656 | ABU07656 Drosophi  |
| 20         | 67.6   | 4.6         | 398    | AA644428 | AA644428 Novel hum |
| 21         | 56.2   | 3.8         | 543    | ABV15942 | ABV15942 Human pro |
| 22         | 55.2   | 3.8         | 470    | ABV45741 | ABV45741 Human pro |
| 23         | 55.2   | 3.8         | 997    | ABV30232 | ABV30232 Human pro |

|    |      |     |        |          |                     |
|----|------|-----|--------|----------|---------------------|
| 24 | 42.4 | 2.9 | 506    | ABQ41711 | ABQ41711 Oligonuc1  |
| 25 | 42.4 | 2.9 | 506    | ABQ41710 | ABQ41710 Oligonuc1  |
| 26 | 42.2 | 2.9 | 2000   | AA271938 | AA271938 Rice gene  |
| 27 | 41.2 | 2.8 | 2000   | ADH71938 | ADH71938 Rice gene  |
| 28 | 39.4 | 2.7 | 23823  | AAK79160 | AAK79160 Human imm  |
| 29 | 39.4 | 2.7 | 23825  | AAK79161 | AAK79161 Human imm  |
| 30 | 39.4 | 2.7 | 110000 | AB279565 | AB279565 CUBD8 and  |
| 31 | 37.8 | 2.6 | 1461   | AA640119 | AA640119 Arabidops  |
| 32 | 37.6 | 2.6 | 605    | AA556911 | AA556911 Human col  |
| 33 | 37.6 | 2.6 | 605    | ABT12333 | ABT12333 Orestes s  |
| 34 | 37.6 | 2.6 | 605    | ACD91627 | ACD91627 Human col  |
| 35 | 37.4 | 2.6 | 2198   | AA97937  | AA97937 Mung bean   |
| 36 | 37.2 | 2.5 | 7498   | ABL32256 | ABL32256 Human imm  |
| 37 | 36.8 | 2.5 | 1578   | ABU11867 | ABU11867 Drosophi   |
| 38 | 36.8 | 2.5 | 4229   | ABU11866 | ABU11866 Drosophi   |
| 39 | 36.6 | 2.5 | 129021 | AAE22295 | AAE22295 BAC conta  |
| 40 | 36.2 | 2.5 | 1814   | AAE22296 | AAE22296 BAC conta  |
| 41 | 36.2 | 2.5 | 3594   | AAE22295 | AAE22295 Murine my  |
| 42 | 35.4 | 2.4 | 1060   | AAE22295 | AAE22295 Murine alp |
| 43 | 35.2 | 2.4 | 1849   | ADA71744 | ADA71744 Rice gene  |
| 44 | 35.2 | 2.4 | 11745  | ABK28332 | ABK28332 DNA trans  |
| 45 | 35   | 2.4 | 241    | AAE17913 | AAE17913 Human bre  |

## ALIGNMENTS

|          |   |                                |
|----------|---|--------------------------------|
| RESULT 1 | ABK53075  | standard; cDNA, 1669 BP.       |
| ID       | ABK53075  |                                |
| AC       | ABK53075  |                                |
| XX       |   |                                |
| DT       | 13-AUG-2002   | (first entry)                  |
| XX       |   |                                |
| DE       | CDNA encoding human 80090 protein.                                    |                                |
| XX       |   |                                |
| XX       | Human; gene; ss; 80090; fucosyl transferase; cancer; Grave's disease; |                                |
| XX       | seven transmembrane receptor; Rho GTPase-activating protein; RhogAP;  |                                |
| XX       | cardiovascular disorder; haematopoietic disorder; brain disorder;     |                                |
| XX       | blood vessel disorder; metabolic disorder; liver disorder; psoriasis; |                                |
| XX       | platelet disorder; leukemia; cardiac hypertrophy; Crohn's disease;    |                                |
| XX       | myocardial infarction; rheumatic fever; multiple sclerosis; asthma;   |                                |
| XX       | rheumatoid arthritis; diabetes mellitus; blood vessel disorder;       |                                |
| XX       | Kaposi sarcoma; Alzheimer's disease; hormonal disorder.               |                                |
| XX       |   |                                |
| OS       | Homo sapiens.   |                                |
| XX       |   |                                |
| FH       | Key   | Location/Qualifiers            |
| FT       | CDs   | 163..1451                      |
| FT       |   | /*tag=a                        |
| XX       |   | /product="Human 80090 protein" |
| XX       |   |                                |
| XX       | WO200232962-A2.   |                                |
| XX       | 25-APR-2002.  |                                |
| XX       | 19-OCT-2001; 2001WO-US048307.   |                                |
| XX       | 20-OCT-2000; 2000US-0241992P.   |                                |
| XX       | 20-OCT-2000; 2000US-0242038P.   |                                |
| XX       | 20-OCT-2000; 2000US-0242040P.   |                                |
| XX       | 23-OCT-2000; 2000US-0242637P.   |                                |
| XX       | (MILL-) MILLENNIUM PHARM INC.   |                                |
| XX       | Gluckmann MA, Meyers R;   |                                |
| XX       | WPI; 2002-463309/49.  |                                |
| XX       | P-PSDB; AAU97908.   |                                |
| XX       | Novel 80090, 52874, 52880, 63497, and 33425 polynucleotides for       |                                |
| XX       | preventing or treating e.g. cancer, cardiovascular, hematopoietic or  |                                |



**This Page Blank (uspto)**

565 ACTACCCATACTTGGAGAGCATTGAAGTCTTGAAGTCACTCCGATACCTAGTTCCTTG 624

```
Search completed: September 13, 2004, 00:41:51
Job time : 3895.02 secs
```

|             |   |                                     |                               |                 |
|-------------|---|-------------------------------------|-------------------------------|-----------------|
| Db          | 1000  | TCCTATTAATAACTGGCAACGATACAAAGTTCA   | TTCTTGCTTTTGAAATGCTATCTGTGAAG | 1059            |
| Qy          | 881   | ACTACATCACTGAGAAAGTTCTGGAGGCCACTGAA | ACTGGGGGCTAGTCCTGTATATTACG    | 940             |
| Db          | 1060  | ATTATATACCTGAAAAACCTCTGGGGGCTCTGAT  | CTGGGAGTGGATACCTGTGCTCTTG     | 1119            |
| Qy          | 941   | GATCCCCCGACATCAACACTGGGCTTCCAGTAA   | CAAAATGCTATTCTTGTATCAGAAAT    | 10000           |
| Db          | 1120  | GTTCTCCACGACTTTTCACCTGGCTTCCACGAA   | CAAGAGTGCACCTTGGATCTAGTT      | 1179            |
| Qy          | 1001  | TTTTCTACCCACGGGAATCGGCAAGTTACATCA   | GCACCTGATTTCTGATGACAGATTGT    | 1060            |
| Db          | 1180  | TTTTACACCCCTTGAGAGCTGGCCCGCTATAT    | CAAAACACTGATTCMAAATGACACGAGAT | 1239            |
| Qy          | 1061  | ATGAGGCCATGTAGATGAGAGCTGAAGGTGAGAT  | CTTAACACGACATCTTGACAG         | 1120            |
| Db          | 1240  | ATGAGGCTTACTGTGAATGTGAAACTGAAAGAC   | ATTTCAAATCCAGGCTGCTTACG       | 1299            |
| Qy          | 1121  | CTCTCAGGAAACGAAATGSGGAGTGCAGACGT    | CAACAGCAATTCATCATGAT          | 1180            |
| Db          | 1300  | CAATGAAGGAACGCAATGGGGAGTGCAGATGT    | CACTCAGAGCAATTTATTGACAT       | 1359            |
| Qy          | 1181  | TTGAGTGTTGGTGTGCACCAAGGTGGGTTAAT    | TACACGCTTCAGGAAAAGGCTTAC      | 1240            |
| Db          | 1360  | TTGAGTGCACTGGTGTGTAAACAGAGTGGGAAA   | CATCAACAGAGAAAGAAAGGATGGC     | 1419            |
| Qy          | 1241  | CACCCAAAAGATGGGAGGCCAAGATATCCCACT   | GAAGTTGCCAGAGCCCAAGTGTTG      | 1300            |
| Db          | 1420  | TGCCCCAGAGTGGAGTGTCTAGGTTAACATCT    | GAATCTCCAAACCTGAGGCTTTG       | 1479            |
| Qy          | 1301  | CTTTCACCACTCCGGAAGTCACTTTGAGCTTT    | GGCAGAGATGTGATTTCCAGCT        | 1360            |
| Db          | 1480  | GTCTCTCTCTTCTT---AATCCCAAGCCAGCT    | CTCCCAAGAGATGTGATAGCAAGTT     | 1539            |
| Qy          | 1361  | TTGAAACATATCAAGAAAGAGCCAGGCACTA     | AGGTGGCTGTGATAGAAATCAAACT     | 1420            |
| Db          | 1536  | TTGACAGATGCCAAGAGAAAGCTTGGCGCTG     | AGGCAAGCTGTGGAAAGAACAGAAAT    | 1599            |
| Qy          | 1421  | TTTCACTTCAGAGATTTTGGGGGCTAGTATT     | CAAGACTGA                     | 1461            |
| Db          | 1596  | TTTCACTTCAGAGATTTTGGATGCTTGTATT     | CAAACTATA                     | 1636            |
| RESULT 15   |   |                                     |                               |                 |
| ACLI15775/c |   |                                     |                               |                 |
| LOCUS       |   |                                     |                               |                 |
| DEFINITION  |   |                                     |                               |                 |
| ACCESSION   | ACLI15775   | 191072 bp                           | DNA linear                    | HTG 24-FEB-2003 |
| VERSION     | ACLI15775   |                                     |                               |                 |
| KEYWORDS    | HTG, HTGS_PHASE1, HTGS_DRAFT.   |                                     |                               |                 |
| SOURCE      | Mus musculus (house mouse)  |                                     |                               |                 |
| ORGANISM    | Mus musculus  |                                     |                               |                 |
| REFERENCE   |   |                                     |                               |                 |
| AUTHORS     | Birren,B., Nusbaum,C. and Lander,E.   |                                     |                               |                 |
| TITLE       | Mus musculus, clone RP23-125P1  |                                     |                               |                 |
| JOURNAL     | Unpublished   |                                     |                               |                 |
| REFERENCE   | 2 (bases 1 to 191072)   |                                     |                               |                 |
| AUTHORS     | Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguski,V.L., Bouckgeater,B., Brown,A., Camarata,J., Campioniano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fairo,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gird,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C., Kama,A., Karitas,A., Kellis,C., Lacroque,K., Lamazares,R., Landers,T., Lehoccky,J., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Margus,N., Matthews,C., McCarthy,M., McEwen,P., McKernan,K., Melittm,J., Menus,L., |                                     |                               |                 |

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 Mihova,T., Mengs,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Punnhang,P., Pierre,N., Pollard,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,U., Rosetti,M., Roy,A., Saitos,R., Schauer,S., Schnupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,U., Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 191072)  
 Birren,B., Nusbaum,C., Lander,E., Abouelell,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B., Camarata,U., Chang,U., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeRellano,K., Faro,S., Diaz,J.S., Dodge,S., Dooley,K., Dorts,U., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,U., Gaidyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,J., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,U., Matthews,C., McCarthy,M., Meldrum,J., Mensus,L., Mihova,T., Mengs,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Punnhang,P., Pierre,N., Rachupka,A., Ramsamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,U., Schauer,S., Schnupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,U., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (24-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Feb 24, 2003 this sequence version replaced gi:28460929.  
 All repeats were identified using RepeatMasker:  
 Smith,A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 -----  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 -----  
 ----- Project Information  
 Center project name: L23395  
 Center clone name: 125\_P 1  
 -----  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 188135 bases at least Q40  
 Consensus quality: 189100 bases at least Q30  
 Consensus quality: 189354 bases at least Q20  
 Insert size: 188000; agarose-EP  
 Insert size: 188972; sum-of-coverage  
 Quality coverage: 11.0 in Q20 bases; agarose-EP  
 Quality coverage: 10.9 in Q20 bases; sum-of-coverage  
 -----  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 12 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1011: contig of 1011 bp in length  
 \* 1012 1111: gap of 100 bp  
 \* 1112 2304: contig of 1193 bp in length  
 \* 2305 2404: gap of 100 bp

Db 1006 TAAGACAAAACCTGGCCCCCATTTGGTATATGTTCACTGGAGCTGGCATCCACCATGACACA 1065  
 QY 701 GGGACAGCTATGTTGGAGCTGATGACTTACATCGAGTTCGATTCCTATGTTGATGTT 760  
 Db 1066 GGGACAGCTATGTTGGAGCTGATGACTTACATCGAGTTCGATTCCTATGTTGATGTT 1125  
 QY 761 TAGAACAACAAGACCTCCCTCAGCAGCTGAAAAATCCACCTCTATGATGTCGATGCT 820  
 Db 1126 TAGAACAACAAGACCTCCCTCAGCAGCTGAAAAATCCACCTCTATGATGTCGATGCT 1185  
 QY 821 TTATAGATCATGTCAGCAGTATAGTTATCTTACCTTTTGAGATGAGATGTTGATG 880  
 Db 1186 TCTACAGATCATGTTGCCAGTATAGTTATCTTACCTTTTGAGATGAGATGTTGATG 1245  
 QY 881 ACTACATCACTGAGAAGTTCTGGAGGCCACTGAACTGGGGGTAGTCCCTGTATATTAG 940  
 Db 1246 ATTACATCACTGAGAAGTTCTGGAGGCCACTGAACTGGGGGTAGTCCCTGTATATTAG 1305  
 QY 941 GATCCCCAGCATCAAGACTGGCTCCAAAGTAAAGTCTATCTGTATAGAT 1000  
 Db 1306 GATCCCCAGCATGCGAGCTGGCTCCAAAGTAAAGTCTATCTGTGTGGAAT 1365  
 QY 1001 TTCTCACCCAGGAGAACTGGCAAGTTACATCAGACGACTGGATTCTGATGACAGATTGT 1060  
 Db 1366 TCTCTACCCAGGAGCTGGAGCTTATTAGAGACTGGATTATGACATGGGTTGT 1425  
 QY 1061 ATGAGGCTTATGTAATGGAATGGAAGTGAAGTCTCTTAACGAGCAGCTTTGACAG 1120  
 Db 1426 ATGAGACCTTATGTAAGTGAAGTGAAGTCTCTTAACGAGCAGCTTTGACAG 1485  
 QY 1121 CTCACGAGGAGCAAAATGGGAGTGCACAGCTCAACAGCAGCAATTACATGATGAT 1180  
 Db 1486 CTCACATGAACCGAGTGGGAGTGCACAGCAGCAATTACAGCAGCAATTACATGATG 1545  
 QY 1181 TTGAGTGTATGTTGTGCACCAAGTGTGGGCTTATATCAAGCTTCAGAAAAAG 1234  
 Db 1546 TTGAGTGTATGTTGTGCACCAAGTGTGGGCTTATATCAAGCTTCAGAAAAAG 1599

RESULT 14  
 GGA535692 1657 bp mRNA linear VRT 23-DEC-2002  
 LOCUS Gallus gallus mRNA for putative alpha3-fucosyltransferase (FUT10  
 DEFINITION gene).  
 ACCESSION AJ535692  
 VERSION AJ535692.1 GI:27368920  
 KEYWORDS alpha3-fucosyltransferase; FUT10 gene.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1  
 REFERENCE  
 AUTHORS Gendeliev, J.J., Martinez-Duncker, I., Oriol, R. and Mollicone, R.  
 TITLE Cloning, expression and genomic organization of a new human  
 JOURNAL alpha3-fucosyltransferase (FUT10)  
 REFERENCE 2 (bases 1 to 1657)  
 AUTHORS Oriol, R.  
 JOURNAL Direct Submission  
 COMMENT Submitted (19-DEC-2002) Oriol R., Glycobiology, INSERM U504, 16 Av.  
 Paul Vaillant-Couturier, 94807 Villejuif, FRANCE  
 FEATURES  
 location/Qualifiers  
 1..1657  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9031"  
 1..1657  
 /gene="FUT10"  
 210..1529  
 /gene="FUT10"  
 /function="adds fucose in alpha1-3 linkage to GlcNAc"

ORIGIN  
 polyA\_signal  
 polyA\_site  
 /codon\_start=1  
 /product="putative alpha3-fucosyltransferase"  
 /protein\_id="CAD59689.1"  
 /db\_xref="GI:27368921"  
 /db\_xref="GOA:O8AMB5"  
 /db\_xref="SPTREMBL:O8AMB5"  
 /translation="WTRRRKRLMASCTCFAPFPVLVTLOVITELNSENKAPAYS  
 LHSGLKPDIERIASQLKNEUYSNREPTDHPILHMSPLTERTSGGQCEDEV  
 PFTINKTYQHSQMTBAPLPYGTDSIDSLPRKDHDWALFHESSPMNYKLFHBA  
 IFLFNHTAFPSHSHLPLTQYLSIEVLRSLRMIPVOMKNSLRKLAPLVVQSDC  
 NAPSDDSYRELMDCHIVDSYGECLHRLDPOHRLMSANDDNFYKILAOYFILA  
 FNAICEDYITEKLMRPLMLGVAPVYFSPSPSIDMLPSNKSALIVSSPSHRLARYI  
 KTLDDNDSEYEVLEWKIGDISNPRILITANKERKGVQVTONYIDTECVKCNRY  
 WENIRREKGMIPQWRSQVNHLCNPLRLSGSPLOIPAPAPSRGCG"  
 1632..1637  
 /gene="FUT10"  
 1653  
 /gene="FUT10"

Query Match 45.0%; Score 657.4; DB 5; Length 1657;  
 Best Local Similarity 68.8%; Pred. No. 9.4e-203;  
 Matches 937; Conservative 0; Mismatches 411; Indels 13; Gaps 2;

Db 101 AGGTATGTTGAGCTGGGGAAGTTGAAAGGAAGGATTTAAAGTTCCAGTTGAG 160  
 Db 289 AGGTTATCACTAACTGGGAAATTTGGGAATTAAGCAGCCAGCAGCTCCAGTTTACACA 348  
 QY 161 ATGACATACAAAAATGGAGGAAGCAGCTTACGATCTTAAATCAATTTCTTAAGAAAG 220  
 Db 349 GTGGCCCACTGAAGCTCATGAGACATGC-----TTCCAGCTTAAAGAAATG 359  
 QY 221 GATTACCTTCAACGAAAAAAGAAATGGAATTTGACAGTATCCCATTAATGCTCTGCT 280  
 Db 400 AACTTATAGCAACTTCAGACAGAGCTGTATAGATCAATCACTACCCATCTGCTCTGCT 459  
 QY 281 GGTCCCGCTGACGGGGAGACTGGAGAGTGAAGCCATGATGAGACAGATGCTTTCT 340  
 Db 460 GGTCTCCCTGACTGAGAGAGAGGAGATGGGGAGTGTGTGAGAGATGTTGTTCT 519  
 QY 341 TCAACATCAACCGAGACTTACTCTCATATCAATGACCAAGCATTTCTCTTATGTA 400  
 Db 520 TTACTATTAAACAAGCGTACAGACAGCTAGATGACAAAGACTTTTGTCTATGTGA 579  
 QY 401 CTGACTTAACTAATATGCTTACCTGCTCGGAAAGCCCATCATGACTGGGCTGTT 460  
 Db 580 CTGACTTCAATATACAGCTTACCTTCTCTCGAAGAAAGCATCATGATTTGGGCTTT 639  
 QY 461 TTCATGAAGAGTCCCGAAAAAACAATTAAAGCTTTTCAATAAACAAGATTAACCTGT 520  
 Db 640 TCCATGAGGAATACCAAAAAACAATTAACCTTTTCAATGAACCGATACCTTAT 639  
 QY 521 TCACTTACAGTCCGCTTCAAGAGAGTATCCCATGCTGCACTAATACCAATACCTTG 580  
 Db 700 TCAACCACTGCAACTTTCAGCGCACTTTCACCTTACGCTTACCACTGAGTACCTTG 759  
 QY 581 AGAGCATGAAGTCCCTGAGTCACTCCGATCACTAGTTCCTTTGACGCAAAAAACAAG 640  
 Db 760 AAGCATGAGAGTCTCTGAGTCTCTGAGACATATGATCCGATGAGTGAAGACAGCC 819  
 QY 641 TTAGAAAAAGACTTCTCCGCTGCTGTATGTAAGTCAAGTGTGAGTGAACCCATGACACA 700  
 Db 820 TGAGGAAGCGGCTTGACCACTTGATGTAAGTGTGAGTGAAGTGTCTCTTGTACC 879  
 QY 701 GGAAGAGCTATGTTGGCAGCTGATGACTTACATCGAGAGTGCATCTCTATGTTGAATGTT 760  
 Db 880 GGGACAGCTATGTAAGTGAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 939  
 QY 761 TAGAACAACAAGACCTCCCTCAGCAGCTGAAAAATCCACCTCTATGATGTCGATGCT 820  
 Db 940 TGACATACAGAGACCTTCTCAGCATTCAGAAATTCATCTGCGCATGATGATGAGAACT 999  
 QY 821 TTATAGATCATGTCAGCAGTATAGTTATCTTACCTTTGAGATGAGATGATGATGATG 880

BC062113 2978 bp mRNA linear ROD 11-DEC-2003  
LOCUS IMAGE:6816067, complete cds.  
DEFINITION BC062113  
ACCESSION BC062113 GI:38328410  
VERSION BC062113.1  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 2978)  
Strasberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Sherman,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buettow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,D., Hsieh,F., DiChicco,L., Marusina,K., Farmer,A., Rubin,G.M., Hong,L., Stopleton,M., Soares,M.B., Bonaldi,M.F., Casavant,T.L., Schecter,I.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Caminchi,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Woley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergran,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Gilmwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalski,U., Smalhus,D.E., Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 2978)  
Strasberg,R.  
Direct Submission  
Submitted (11-NOV-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
REMARK COMMENT  
Contact: MGC help desk  
Email: [cgabbe@mail.nih.gov](mailto:cgabbe@mail.nih.gov)  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
CDNA Library Preparation: M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLB)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Gun, Leticia Hsieh, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Diane Smalhus, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLB at: <http://image.llnl.gov>  
Series: IRAX Plate: 132 Row: h Column: 16  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency OR analysis.  
FEATURES  
source location/Qualifiers  
1..2978  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="MGC:69807 IMAGE:6816067"  
/tissue\_type="Brain, mouse, 13.5, 14.5, 16.5, 17.5 dpc"

gene  
/clone\_lib="NIH\_BMAP\_Fw0"  
/lab\_host="DH10B"  
/note="Vector: pYX-Asc"  
1..2978  
/gene="Fut10"  
/note="synonym: MGC40919"  
/db\_xref="LocusID:171167"  
/db\_xref="MGI:2384748"  
387..1697  
/codon\_start=1  
/product="fucosyltransferase 10"  
/protein\_id="AAH62113.1"  
/db\_xref="GI:38328410"  
/db\_xref="LocusID:171167"  
/translation="WRFQRKLILASLCVATRVFLVNTLVYVLEKFEKKIKDSN  
VDGHRDVGEBKLEPPEKEALALAGRTKVDASVIVVMSVLTETRLQCGA  
DACEFTINRFOHHPMTRAFLEFYGDFFNIDSLPERAHDMALPHESPKNTKLF  
KPYITLFRNTAFSRHSLPLTQYLEGVDYKSLRYVLPQAKNNLQKLAPLYVQ  
SCDDPSDSDSVRLMAYIEVDSYGECCJONDLPOOLKNPASMDADLFYVIAQYK  
ILAFENAVDDYITTEKMPKLTGLVPYVGSPTIADMLPSRSAILVSESHPELA  
SIFRLIDYDGLYETVEKLLKGSINQSLTLNERNRWGVDINQDNYIDFECMC  
RRVWANSRLQEQVSEWKSQGMHPSLCVAVLFLMWLPAITGLYS"  
633..1574  
/note="Glyco. transf 10; Region: Glycosyltransferase family  
10 (fucosyltransferase). This family of  
Fucosyltransferases are the enzymes transferring fucose  
from GDP-Fucose to GlcNAc in an alpha,3 linkage. This  
family is know as glycosyltransferase family 10"  
/db\_xref="COD:p1am0852"  
ORIGIN  
Query Match 54.7%; Score 799.6; DB 10; Length 2978;  
Best local similarity 81.6%; Pred. No. 3.2e-249;  
Matches 925; Conservative 0; Mismatches 209; Indels 0; Gaps 0;  
misc\_feature  
101 AGGTGATGTTGAGCTGGGGAAGTTGAAAGAGAGATTAAAGTTCAAGTTGCAG 160  
466 AGGAGTGGTTGAGCTTGGGAAATTGAAAGAGAGAGCTTAAAGACTCAATGTGCAG 525  
161 ATGGACATTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 220  
526 AGGACATTAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 585  
221 GATTGACCTTCAACAG 280  
586 CACGGGCTTGGCTGGAG 645  
281 GGTCCCGCTGACGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 340  
646 GGTCCCACTGACGAG 705  
341 TCACCATCAACCGGACCTACCTCCATCATCATGACCAAGCATTCCTTCTATGTA 400  
706 TTACCATCAACAG 765  
401 CTGACTTAACTAGATGAGTACCTTCTGCTGGGAAAGGCCATCATGACTGGGCTTT 460  
766 CTGACTTAACTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 825  
461 TTCAATGAAGAGTCCCGCAAAACATATTAAGCTCTTCTTAACAGAGATTACCTGT 520  
826 TCATGAAGAGTCCCGCAAAACATATTAAGCTCTTCTTAACAGAGATTACCTGT 585  
521 TCAACTACCTGACGAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 580  
886 TCAACATACGAGGACATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 945  
581 AGAGATTGAAGTCCCTGAAGTCCGATACCTAGTCTTCTTGGAGTCCCAAAACAGC 640  
946 AAGGCTGAGAGTCTTAAGTCACTGATACCTAGTCTTCTTTCAGAGCTTAAGAACAC 1005  
641 TTAGAAAAAGACTGCTCGGTGTGTATGATAGTCACTGAGCCATCAGATCAGACA 700

## REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Akhter, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brixley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,  
Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantipop, S., Thomas, P.J., Touchman, J.W.,  
Tsurgoun, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.jhl.gov>  
Series: IRAX Plate: 65 Row: h Column: 14  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis.

## FEATURES

## Source

## Location/Qualifiers

1..2874

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="MGC:40819 IMAGE:5368333"

/tissue\_type="Eye, retina, mouse strain C57BL/6"

/clone\_lib="NIH\_MGC\_94"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

1..2874

/gene="Fut10"

/note="synonym: MGC40819"

/db\_xref="locusid:171167"

/db\_xref="MGI:2584748"

285..1595

/codon\_start=1

/product="fucosyltransferase 10"

/protein\_id="AAH30474.1"

/db\_xref="MGI:2098074"

/db\_xref="locusid:171167"

/translations="VVRFORRLKSLASCLCTVATVFLMTLOVVELKFEKKLKJSDN  
VQGRHVDVEGEPHLEPPPEKALALAGRTKVDASYSIVLWSEPLTETRLQCCOA  
DACFTFNIRFOHNPWTRAFLEFGDFNIDSJLPRHSHMALFHESPKNNKYLKH  
KPVITLFNHTATSRSHSLPLTQYLEGVILKSLRVLPLOAKNNRQKLAPLYQV  
SDCDPSRDSYVRELMAYIEVDSYGECLONRDLPOOLKNPSMDADAFYVIOYE  
ILAFENAVDDDYTEKRWELKLGIVPYVYGSPTIADLPENRSALIVSESHRELA  
SFIRLDYDGGATYETVWKIKGKISNQLPLALERKMGVODINODNYIDSFECMVC  
RRVWANSRLQBOVSEWKSQGMHPSLCVTLVFLMLWLPATGLYS"

447..1514

/note="Glyco. transf 10; Region: Glycosyltransferase family  
10 (fucosyltransferase). This family of  
Fucosyltransferases are the enzymes transferring fucose  
from GDP-Fucose to GlcNAc in an alpha,3 linkage. This  
family is know as glycosyltransferase family 10"  
/db\_xref="CD: Pfam00852"

## ORIGIN

Query Match 54.7%; Score 799.6; DB 10; Length 2874;  
Best Local Similarity 81.6%; Pred. No. 3.2e-249;  
Matches 925; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 101 AGGTCAATGTTGAGCTGGGGAAGTTGAAAGAGAGAGTTTAAAGTTCCAGTTTGCAAG 160  
Db 364 AGGTAGGCTTGAGCTGGGAAATTGAAAGAGAAAGAGTTTAAAGTCCAGTGTGCAAG 423  
QY 161 ATGACATACAAATAAGAGAGACCTTAACTTATTCATTTCTTAAGAAAGAG 220

424 ACGACATAGAGATGTGAGAGGAGCCCTAAACACCTGGAGCCCTTCCCTGAAAAAGAG 483  
QY 221 GATTGACCTTCAACAGAGAAAGAAATGGGAATTGGACAGCTAACCCATTATGCTTCGT 280  
Db 484 CACTGGCTTGTGCTGGAGAGACAAAGTGAATGCTGATGACCTAACCTGCTGTGCT 543  
QY 281 GATCCCGCTGACGGGGAGAGTGGAGGTTAGCCAAATGTGAGACAGATGCTGTCTTCT 340  
Db 544 GATCCCACTGACTGAGAGAGACTGGAGGCTTGGCCAGTGTGGGCAAGTGTCTTCTCT 603  
QY 341 TCACCATCAACGGGAGACTTCTTCATCATCATCATGACCAAGACATTCCTCTTATGATA 400  
Db 604 TTACCATCAACAGAGACTTTCAGCATCATCATGACCAAGGAGCTTCTTATTCAGGATA 663  
QY 401 CTGACTTTAATAGATAGTATAGTATAGTCTCGGAAAGCCCATGACTGAGGCTGTTT 460  
Db 664 CTGACTTTAATAGATAGTATAGTATAGTCTCGGAAAGCCCATGACTGAGGCTGTTT 723  
QY 461 TTGATGAAAGTGTCCCGAAAAACAATTATAGCTCTTTCATTAACAGTATACCTTGT 520  
Db 724 TCCATGAAAGTGTCCCGAAAAACAATTATAGCTCTTTCATTAACAGTATACCTTGT 783  
QY 521 TCACCTACACCTGACAGCTTACAGAGAGCTTCCCATTCGACCTAACCTAACCTTGG 580  
Db 784 TCACCTACACCTGACAGCTTACAGAGAGCTTCCCATTCGACCTAACCTAACCTTGG 843  
QY 581 AGAGCATTAAGTCTGAAAGTCACTCCGATACCTTATAGCTCTTTCGATCCAAAAAGAG 640  
Db 844 AAGGCGTGAAGTCTGAAAGTCACTCCGATACCTTATAGCTCTTTCGATCCAAAAAGAG 903  
QY 641 TTGAAAAAGACTTGTCTCCGCTGTGTATGTATAGTCAAGTGTGACCCATCATCAGACA 700  
Db 904 TTAGCAAAAAAGTGTCTCCGCTGTGTATGTATAGTCAAGTGTGACCCATCATCAGACA 963  
QY 701 GGGACAGTATGTGCGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 760  
Db 964 GGGACAGTATGTGCGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1023  
QY 761 TACGAAACAAAGCTCTCTCTGACGAGCTGAAAAATCCAGCTCTATGAGTCCGATGCT 820  
Db 1024 TACGAAACAAAGCTCTCTCTGACGAGCTGAAAAATCCAGCTCTATGAGTCCGATGCT 1083  
QY 821 TTATATGATCATGACAGATTAAGTATATGATGATGATGATGATGATGATGATGATG 880  
Db 1084 TCTACAGATCATTTGCGAGATTAAGTATATGATGATGATGATGATGATGATGATGATG 1143  
QY 881 ACTACATCATGAGAAAGTTCTGAGGCGAAGTGAAGCTGAGGAGTCTGATATATTACG 940  
Db 1144 ATTACATCATGAGAAAGTTCTGAGGCGAAGTGAAGCTGAGGAGTCTGATATATTACG 1203  
QY 941 GATCCCGGAGATCAAGAGCTGCTTCCAAATGAAAGAGCTATCTGATACAGAAAT 1000  
Db 1204 GATCCCGGAGATCAAGAGCTGCTTCCAAATGAAAGAGCTATCTGATACAGAAAT 1263  
QY 1001 TTTCTACCCCGAGAGAGTGGCAAGTTACATCAGACGACTGATTCGATGACAGATTGT 1060  
Db 1264 TTTCTACCCCGAGAGAGTGGCAAGTTACATCAGACGACTGATTCGATGACAGATTGT 1323  
QY 1061 ATGAGGCTATGTATAGAAAGAGCTGAAAGGTGAGATCTTACACAGCACTTGTGACAG 1120  
Db 1324 ATGAGGCTATGTATAGAAAGAGCTGAAAGGTGAGATCTTACACAGCACTTGTGACAG 1383  
QY 1121 CTCTCAGGAAACGAAATGAGGAGTGCAGAGCTCAACAGAGCAATTAATCATGATCAT 1180  
Db 1384 CTCTCAGGAAACGAAATGAGGAGTGCAGAGCTCAACAGAGCAATTAATCATGATCAT 1443  
QY 1181 TTGAGTGTATGTGTGACCAAGGTGTGGCTTAAATTCAGGCTTCAGAAAAAG 1234  
Db 1444 TTGAGTGTATGTGTGACCAAGGTGTGGCTTAAATTCAGGCTTCAGAAAAAG 1497

```

repeat_region      /rpf_family="AlusX"
                    8228..8249
                    /rpf_family="AT_rich"
repeat_region      complement(8522..8823)
                    /rpf_family="AlusX"
repeat_region      8824..8899
                    /rpf_family="MIR"
repeat_region      9039..9236
                    /rpf_family="MIR"
                    complement(9469..9776)
repeat_region      /rpf_family="AlubD"
                    complement(9802..9904)
repeat_region      /rpf_family="L1MC5"
                    9998..10295
repeat_region      /rpf_family="AlusG"
                    complement(10874..10935)
repeat_region      /rpf_family="L1MC5"
                    10942..11275
repeat_region      /rpf_family="Charlie1b"
                    complement(11276..11577)
repeat_region      /rpf_family="AlusX"
                    11578..11708
repeat_region      /rpf_family="Charlie1b"
                    complement(11733..12029)
repeat_region      /rpf_family="AlusG"
                    complement(12082..12251)
repeat_region      /rpf_family="FRAM"
                    12335..12629
repeat_region      /rpf_family="AlusX"
                    complement(12630..12743)
repeat_region      /rpf_family="L1MB6"
                    complement(12744..13041)
repeat_region      /rpf_family="AlusG"
                    complement(13042..13612)
repeat_region      /rpf_family="L1MB6"
                    13613..13668
repeat_region      /rpf_family="Charlie1b"

```

Query Match 57.4%; Score 838.8; DB 9; Length 185872;  
 Best Local Similarity 99.2%; Pred. No. 9.9e-262;  
 Matches 843; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

355 TTCCTCTCTAGTGTACGACTTAACATGATAGTACCTTCTGCTGGGAAAGCCAT 444
144150 TTTCTTCTCTAGTGTACCTTTAATATGATAGTACCTTCTGCTGGGAAAGCCAT 144209
445 CATGACTGGGCTGTTTTCATGAGAGTCCCGGAAAAACATTAATAGCTCTTTCTATAA 504
144210 CATGACTGGGCTGTTTTCATGAGAGTCCCGGAAAAACATTAATAGCTCTTTCTATAA 144269
505 CGAGGATTAAGTCTGTTCACTACACGCAAGTTCAGCAGGATTCCTCCACTTGCACATA 564
144270 CGAGGATTAAGTCTGTTCACTACACGCAAGTTCAGCAGGATTCCTCCACTTGCACATA 144329
565 ACTACCCAAATCTTGAGAGCATGAGTCTGTAAGTCACTCCGATACCTAGTCCCTTGG 624
144330 ACTACCCAAATCTTGAGAGCATGAGTCTGTAAGTCACTCCGATACCTAGTCCCTTGG 144389
625 CAGTCCAAAAACAAGCTTAGAAAAAGACTTGCCTCGCTGTAATGATACAGTCAAGCTGT 684
144390 CAGTCCAAAAACAAGCTTAGAAAAAGACTTGCCTCGCTGTAATGATACAGTCAAGCTGT 144449
685 GACCCACCATCAGACAGGAGCAGTATGTTGGAGGTGATGCTTACATCGAGGTCGAT 744
144450 GACCCACCATCAGACAGGAGCAGTATGTTGGAGGTGATGCTTACATCGAGGTCGAT 144509
745 TCCTATGATGATGTTTACGAAACAAAGACCTCTCAGCAGCTGAAAAATCCAGCTCT 804
144510 TCCTATGATGATGTTTACGAAACAAAGACCTCTCAGCAGCTGAAAAATCCAGCTCT 144569
805 ATGGAATGCCGATGCTTTTATAGGATATGACAGATTAAGTTTATCTTACTTTTGG 864
144570 ATGGAATGCCGATGCTTTTATAGGATATGACAGATTAAGTTTATCTTACTTTTGG 144629

```

```

QY 865 AATGCACTTTGTATGATCACTACATCACTGAGAACTTCTGAGGCCCATGAAATCGGGGTA 924
144630 AATGCACTTTGTATGATCACTACATCACTGAGAACTTCTGAGGCCCATGAAATCGGGGTA 144689
QY 925 GTCCCTGTATATTACGAGATCCCCAGCATCATCAGACTGGCTCCCAAGTAAACAAAGTGT 984
144690 GTCCCTGTATATTACGAGATCCCCAGCATCATCAGACTGGCTCCCAAGTAAACAAAGTGT 144749
QY 985 ATTCTGTATCAGAAATTTTCTCAACCCAGGGAACGTGCAAGTTTATCATCAGACGATGAT 1044
144750 ATTCTGTATCAGAAATTTTCTCAACCCAGGGAACGTGCAAGTTTATCATCAGACGATGAT 144809
QY 1045 TCTGATGACAGATTGTATGAGGCTTGTATGATGGAAGCTGAAGGGAGATCTCTAAC 1104
144810 TCTGATGACAGATTGTATGAGGCTTGTATGATGGAAGCTGAAGGGAGATCTCTAAC 144869
Db 144870 CAGCGACTTCTGACAGCTCTCAGAGGAAACGGAATGGGGAGTCAAGACGTCAACAGGAC 144929
QY 1105 CAGCGACTTCTGACAGCTCTCAGAGGAAACGGAATGGGGAGTCAAGACGTCAACAGGAC 1164
144930 CAGCGACTTCTGACAGCTCTCAGAGGAAACGGAATGGGGAGTCAAGACGTCAACAGGAC 144989
QY 1165 AATTACATGATGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1224
144990 AATTACATGATGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 144999
Db 1225 CAGGAAAGG 1234
144990 CAGGAAAGG 144999

```

RESULT 12  
 BC030474 2874 bp mRNA linear ROD 06-OCT-2003  
 LOCUS Mus musculus fucoseyltransferase 10, mRNA (cDNA clone MGC:40819  
 IMAGE:5568333), complete cds.  
 ACCESSION BC030474  
 VERSION BC030474.1 GI:20988073  
 KEYWORDS WGC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Berger, J.G.,  
 Klusner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buelow, K.H., Scheefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Shapiro, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Rhee, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullany, S.J., Bosak, S.A., McSwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hally, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Hellon, E., Kettleman, M., Madan, A., Young, A.C., Shcherbakov, Y.,  
 Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butlerfield, Y.S., Krzywinski, M.I., Skalka, J., Smalls, D.E.,  
 Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 MEDLINE 22388257  
 PUBMED 12477932  
 REFERENCE 2 (bases 1 to 2874)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-MAY-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gorette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., Lakoque, K., Lamatares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., MacDonald, P., Marquis, N., Matthews, C., McCarthy, M., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhahng, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (01-APR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 185872)

REFERENCE  
AUTHORS  
Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barrera, N., Bastien, V., Bloom, T., Boguski, L., Bouckgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, J., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Ton, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhahng, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
AUTHORS  
REFERENCE  
Submitted (25-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 185872)

Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barrera, N., Bastien, V., Bloom, T., Boguski, L., Bouckgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, J., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Ton, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhahng, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT  
Submitted (05-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Nov 5, 2002 this sequence version replaced gi:4371492.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: <http://www-seq.wi.mit.edu>

| FEATURES      | source    | location/Qualifiers   |
|---------------|-----------|---|
| repeat_region | 1..185872 | <p>Contact: sequence_submissions@genome.wi.mit.edu</p> <p>----- Project Information</p> <p>Center project name: L13069</p> <p>Center clone name: 359_B_20</p> <p>-----</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:9606"</p> <p>/chromosome="8"</p> <p>/map="8"</p> <p>/clone="RP11-359B20"</p> <p>/clone_lib="RP11-11 Human Male BAC"</p> <p>/complement(1..273)</p> <p>/rpt_family="MER50"</p> <p>complement(1253..503)</p> <p>/rpt_family="MER50"</p> <p>521..549</p> <p>/rpt_family="(A)n"</p> <p>561..812</p> <p>/rpt_family="AluSc"</p> <p>complement(813..918)</p> <p>/rpt_family="L2"</p> <p>924..1036</p> <p>/rpt_family="FLAM C"</p> <p>complement(1132..1628)</p> <p>/rpt_family="MLT2B2"</p> <p>complement(1694..1981)</p> <p>/rpt_family="AluIo"</p> <p>complement(2063..2368)</p> <p>/rpt_family="AluSc"</p> <p>2398..2479</p> <p>/rpt_family="L1MC4a"</p> <p>2480..2740</p> <p>/rpt_family="AluSc"</p> <p>2741..3183</p> <p>/rpt_family="L1MC4a"</p> <p>3202..3323</p> <p>/rpt_family="FLAM C"</p> <p>3324..3352</p> <p>/rpt_family="(TAAA)n"</p> <p>3398..3636</p> <p>/rpt_family="L1MC5"</p> <p>3775..3857</p> <p>/rpt_family="L1MC5"</p> <p>3862..4125</p> <p>/rpt_family="AluSc"</p> <p>4127..4439</p> <p>/rpt_family="AluSc"</p> <p>4646..4939</p> <p>/rpt_family="AluIo"</p> <p>4983..5004</p> <p>/rpt_family="AT-rich"</p> <p>5024..5046</p> <p>/rpt_family="(T)n"</p> <p>complement(5071..5198)</p> <p>/rpt_family="FLAM C"</p> <p>complement(5190..5287)</p> <p>/rpt_family="AluSc"</p> <p>5345..5658</p> <p>/rpt_family="AluSc"</p> <p>5908..6196</p> <p>/rpt_family="AluSc"</p> <p>6290..6426</p> <p>/rpt_family="AluIo"</p> <p>6427..6431</p> <p>/rpt_family="AT-rich"</p> <p>complement(6592..6893)</p> <p>/rpt_family="AluSc"</p> <p>complement(7502..7782)</p> <p>/rpt_family="AluSc"</p> <p>complement(7840..8149)</p> |



```

repeat_region      complement(984, .1374)
                    /rpl_family="MLTIF"
repeat_region      complement(1401, .1576)
                    /rpl_family="MLTIF-int"
repeat_region      complement(1940, .2100)
                    /rpl_family="MLTIFH1"
repeat_region      complement(2427
                    /rpl_family="L2"
repeat_region      complement(4278, .4692)
                    /rpl_family="MLTIF1"
repeat_region      complement(4818, .4937
                    /rpl_family="L2"
repeat_region      complement(4941, .5075
                    /rpl_family="MIR"
repeat_region      complement(6708, .6797
                    /rpl_family="L2"
repeat_region      complement(6892, .6955)
                    /rpl_family="L2"
repeat_region      complement(7144, .7447)
                    /rpl_family="AluYo"
repeat_region      complement(7745, .7854
                    /rpl_family="L2"
repeat_region      complement(7907, .7982)
                    /rpl_family="MIR"
repeat_region      complement(8737, .8773
                    /rpl_family="CAAAA)n"
repeat_region      complement(9080, .9150)
                    /rpl_family="MLTIF"
repeat_region      complement(10454, .10537)
                    /rpl_family="L2"
repeat_region      complement(10538, .10643
                    /rpl_family="AluSg"
repeat_region      complement(10844, .10894)
                    /rpl_family="L2"
repeat_region      complement(10971, .11034
                    /rpl_family="MIR"
repeat_region      complement(11247, .11379
                    /rpl_family="L2"
repeat_region      complement(11381, .11425
                    /rpl_family="AT-rich"
repeat_region      complement(11434, .11721
                    /rpl_family="AluSg"
repeat_region      complement(11916, .12037
                    /rpl_family="AluYo"
repeat_region      complement(12038, .12350)
                    /rpl_family="AluY"
repeat_region      complement(12351, .12539
                    /rpl_family="AluYo"
repeat_region      complement(12540, .12559
                    /rpl_family="GGA)n"
repeat_region      complement(12578, .12689
                    /rpl_family="AluSx"
repeat_region      complement(12903, .12923
                    /rpl_family="(GGGA)n"
repeat_region      complement(13576, .13750)
                    /rpl_family="L2"
repeat_region      complement(13954, .14317)
                    /rpl_family="L2"
repeat_region      complement(14533, .14599)
                    /rpl_family="VERSA"
repeat_region      complement(14600, .14691)

Query Match      57.4%; Score 838.8; DB 9; Length 104120;
Best Local Similarity 99.2%; Pred. No. 9e-262;
Matches 843; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 505 CCAAGTATTAACCTTGTTCACATCACTGCGACAGCTTCAGAGGCATTTCCACATTGGCACA 564
DB 70067 CCAGTATACCTTGTTCACATCACTGCGACAGCTTCAGAGGCATTTCCACATTGGCACA 70008
QY 565 ACTACCAATACCTTGGAGAGCATTTGAAGTCTGAAGTCACTCCGATACCTAGTCTTTG 624
DB 70007 ACTACCAATACCTTGGAGAGCATTTGAAGTCTCGAAGTCACTCCGATACCTAGTCTTTG 69948
QY 625 CAGTCCAAAAACAAGCTTGAAAAAACAATCTTCTCGAGTGTGTATGTACAGTCAAGTCT 684
DB 69947 CAGTCCAAAAACAAGCTTGAAAAAACAATCTTCTCGAGTGTGTATGTACAGTCAAGTCT 69888
QY 685 GACCCACATCAGACAGGACAGCTATGTTGGAGCTGATGATCACTTAATCAATCAGAGTGCAT 744
DB 69887 GACCCACATCAGACAGGACAGCTATGTTGGAGCTGATGATCACTTAATCAATCAGAGTGCAT 69828
QY 745 TCCATATGTGAATGTTTACGAAAACAAGACCTCCCTCAGACAGCTGAAAAATCCAGCTCT 804
DB 69827 TCCATATGTGAATGTTTACGAAAACAAGACCTCCCTCAGACAGCTGAAAAATCCAGCTCT 69768
QY 805 ATGATGCGGATGCTTTATATGATCATTTGACAGATTAATGATTAATCCTAGCTTTGAG 864
DB 69767 ATGATGCGGATGCTTTATATGATCATTTGACAGATTAATGATTAATCCTAGCTTTGAG 69708
QY 865 AATGCAATTTGTATGATGATCACTACATCACTGAGAAATTTCTGGAGGCACTGAACTGGGGTA 924
DB 69707 AATGCAATTTGTATGATGATCACTACATCACTGAGAAATTTCTGGAGGCACTGAACTGGGGTA 69648
QY 925 GTCCCTGTATATTAAGGATCCCCCAGCATCAAGACTGGCTTCCAAAGTAACAAGTCT 984
DB 69647 GTCCCTGTATATTAAGGATCCCCCAGCATCAAGACTGGCTTCCAAAGTAACAAGTCT 69588
QY 985 ATTCTGTATACAGAAATTTTCTACCCCAAGGAAGTGGCAATTAATCAAGACAGTGTAT 1044
DB 69587 ATTCTGTATACAGAAATTTTCTACCCCAAGGAAGTGGCAATTAATCAAGACAGTGTAT 69528
QY 1045 TCTGATGACAGATTTGATGAGGCTTATGAAATGGAAGCTTAAGGGTGAATCTCTTAC 1104
DB 69527 TCTGATGACAGATTTGATGAGGCTTATGAAATGGAAGCTTAAGGGTGAATCTCTTAC 69468
QY 1105 CAGGACCTTCTGACAGCTCTCAAGGGAACGGAATGAGGAGTCAAGACGTCACACAGGAC 1164
DB 69467 CAGGACCTTCTGACAGCTCTCAAGGGAACGGAATGAGGAGTCAAGACGTCACACAGGAC 69408
QY 1165 AATTACATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1224
DB 69407 AATTACATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 69348
QY 1225 CAGGAAAAAG 1234
DB 69347 CAGGAAAAAG 69338

RESULT 11
AC091144 165872 bp DNA linear PRI 05-NOV-2002
LOCUS Homo sapiens chromosome 8, clone RP11-359B20, complete sequence.
DEFINITION AC091144
ACCESSION AC091144
VERSION AC091144.12 GI:24580467
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 165872)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguski,K., Brown,A., Brockway,E., Brown,A.,
Camarata,J., Campolongo,A., Chang,J., Choquet,Y., Colangelo,M.,

```

Cy 1421 TTTCATCTCAGAGTTTGGGGCTGTATTCAGGACTGA 1461  
 Db 1349 TTTCATCTCAGAGTTTGGGGCTGTATTCAGGACTGA 1389  
 RESULT 10  
 LOCUS AC067838/c 104120 bp DNA linear PRI 21-MAR-2003  
 DEFINITION Homo sapiens chromosome 8, clone RP11-722E23, complete sequence.  
 ACCESSION AC067838  
 VERSION AC067838.18 GI:29135646  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 104120)  
 AUTHORS Birren, B., Nusbaum, C. and Lander, E.  
 TITLE Homo sapiens chromosome 8, clone RP11-722E23  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 104120)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,  
 Boguslavskiy, L., Bouckgalter, B., Brown, A., Burkett, G.,  
 Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J., S.,  
 Dodge, S., Domingo, M., Doyle, K., Ferreira, P., FitzHugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heistord, A., Horton, L.,  
 Howland, J., C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A.,  
 Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,  
 Melchior, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morcov, J.,  
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Testfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE 3 (bases 1 to 104120)  
 AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
 Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B.,  
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
 Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J., S., Dodge, S.,  
 Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,  
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,  
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,  
 Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
 Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C.,  
 McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Plunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,  
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,  
 Testfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-OCT-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE 4 (bases 1 to 104120)  
 AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
 Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B.,  
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
 Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J., S., Dodge, S.,  
 Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,

Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J.,  
 Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T.,  
 Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
 Peterson, K., Plunkhang, P., Pierre, N., Raymond, C., Retta, R.,  
 Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R.,  
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Talamas, J., Testfaye, S., Theodore, J., Topham, K.,  
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-NOV-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE 5 (bases 1 to 104120)  
 AUTHORS Birren, B., Nusbaum, C., Lander, E., Aboueleil, A., Allen, N.,  
 Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,  
 Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y.,  
 Collymore, A., Cook, A., Cooke, P., Corum, B., Deatellano, K.,  
 Diaz, J., S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S.,  
 Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,  
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,  
 MacDonald, P., Major, J., Manning, J., Mlenga, V., Murphy, T.,  
 Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Norman, C. H.,  
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Plunkhang, P., Pierre, N.,  
 Rachtupka, A., Ramsamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
 Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-MAR-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Mar 21, 2003 this sequence version replaced gi:24496795.  
 All repeats were identified using RepeatMasker:  
 Smith, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L9405  
 Center clone name: 722\_E\_23  
 -----  
 Only the first 104,1 Kilobases of the middle contig is being  
 submitted.  
 This clone is completely overlapped by number AC09114 [WICR  
 Project L9369] and AC104027 [WICR Project L21993].  
 Location/Qualifiers  
 1. 104120  
 /organism="Homo sapiens"  
 /mol\_type="Genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="8"  
 /map="8"  
 /clone="RP11-722E23"  
 /clone\_lib="RP11 Human Male BAC"  
 9. 142  
 /rpt\_family="(TA)n"  
 repeat\_region complement(323..378)  
 /rpt\_family="(T2)"  
 repeat\_region complement(865..923)  
 /rpt\_family="(MT)n"

QY 1203 GGTGTGGCTAATATCAAGCTTCAGAAAAG 1234  
DB 1554 GGTGTGGCTAATATCAAGCTTCAGAAAAG 1585

RESULT 9  
AY344579 1389 bp mRNA linear MAM 03-AUG-2003  
LOCUS AY344579  
DEFINITION Bos taurus putative alpha 1,3-fucosyltransferase (fuc10) mRNA,  
complete cds.  
ACCESSION AY344579  
VERSION AY344579.1 GI:33303527  
KEYWORDS  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 1389)  
Germot,A., Dupuy,F., Lorioi,C., Julien,R. and Maftah,A.  
Genomic organization and expression profile of putative alpha  
1,3-fucosyltransferase genes, fuc10 and fuc11, in Bos taurus  
Unpublished  
2 (bases 1 to 1389)  
Germot,A.  
Direct Submission  
Submitted (17-JUL-2003) EA 3176 Glycobiologie et Biotechnologie,  
Faculte des Sciences et Techniques, 123, Avenue A. Thomas, Limoges  
87060, France

FEATURES  
source location/Qualifiers  
1..1389  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
1..1389  
/gene="fuc10"  
1..1389  
/gene="fuc10"  
/codon\_start=1  
/product="putative alpha 1,3-fucosyltransferase"  
/protein\_id="AA02331.1"  
/translation="MATVFLILTLQVVELGKPEGKKFNNSHLKDGAOMEAPLHL  
PFNRREGILNKRKTLADSPIMLWMSPLTGTGTCGCGADACFTTINRYLHHR  
TKALFLYGTDSIDSLPLPKAHMDMLFHEBSPKNYLFLPOVLTLENYATSRH  
SHLTLTOYLEGTEVLTLRLHLVPLRSKHLRKSLLPYVYSDCPSPDRSYREL  
MTYIEVDSKCELRNKLPLPOLSNPASMDDGFRLIAYKFLIAPENAVCDYITBK  
LMLRLKGVVPEVYIGSPSIADWLPSNSALVSEFSHPBLASYIALDRDRYQAY  
IEMRLKGEISQRLTLALRRKMGVQDVKQNDYIDAPECMVCKEWDNIRLDEKGLP  
KRMQADVTHLSCEPPTVFASFPLVPRRRSLREMWIPSFQSKKEAALFMLVDENQNF  
STGEFMALVFED"

ORIGIN

Query Match 69.4%; Score 1013.8; DB 4; Length 1389;  
Best Local Similarity 84.6%; Pred. No. 0;  
Matches 1151; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 101 AGGTCATGTTGACCTGGGGAAGTTGAAAGAGAGATTAAAGTCCAGTTTCAAG 160  
DB 32 AGGTGCGTGTGACCTGGGAAATTTGAAGCAAGAGTTTAAATTCATTAAAG 91  
QY 161 ATGACATACAAAATGAGAGAGACACCTTAGCATCTTAATTATTCTTAAAGAAAG 220  
DB 92 ATGACATAGCAAAATGAGAGAGACACCTTAGCATCTTAATTATTCTTAAAGAAAG 151  
QY 221 GATTGACCTTCAACAGGAAAGAAATGAGAAATGAGACACTACCCCATATGCTTGT 280  
DB 152 GACTGACCTTCAACAGGAAAGAAATGAGAAATGAGACACTTCCCATATGCTTGT 211  
QY 281 GGTCCCCCTGACCGGGGAGACTGGAGGTTAGGCCATGTGAGAGAGATGCTTGT 340  
DB 212 GGTCCCCCTGACCGGGGAAACCGAAGGCTCGGCGAGTGTGGGCGGAGTGTGCTT 271

QY 341 TCACATCAACCGGACCTACCTTCATCATACATGACCAAGACATTCCTTTATAGTA 400  
DB 272 TCACATCAACCGGACCTACCTTCATCATACATGACCAAGACATTCCTTTATAGTA 331  
QY 401 CTGACTTTAATAGATAGTTTACCTCTGCTGGAAGCCCATCATGACTGGCTGTT 460  
DB 332 CTGACTTTAATAGATAGTTTACCTCTGCTGGAAGCCCATCATGACTGGCTGTT 391  
QY 461 TTCAATGAGAGTCCCGGAAACCAATTAATAGCTCTTTATTAACAGAGATTAAGT 520  
DB 392 TCACGAAAGATCTCCAAAATAATTAAGCTCTTTACCAAGCAGATCACCTTGT 451  
QY 521 TCAACTACCTGCAAGTTGACGAGCATCTCCACTTGGCACTTAATCCCAATCTTGG 580  
DB 452 TCAACTACCTGCAAGTTGACGAGCATCTCCACTTGGCACTTAATCCCAATCTTGG 511  
QY 581 AGAGCATTAAGTCTGGAAGTCACTCCGATACCTAGTCTTTGGAGTCCAAAACAGC 640  
DB 512 AGGCAAGGAAAGTCTGAGCTACTCCGACCTGTTCCCTTGGGTCGCAAAACAC 571  
QY 641 TTGAAAAAGACTTGTCTCGCTGGTGTATGTACAGTCAAGCTGTGACCCACATGACA 700  
DB 572 TCCGAAAGAGCTTCGCGCCGCTGTGTATAGTACAGTCCAGCTGTGACCCGCTGACA 631  
QY 701 GGAACAGCTATGTTCCGAGCTGATGACTTACATGAGAGTGCATTCATGAGTATGT 760  
DB 632 GGAACAGTTCAGTTGAGAGCTGATGACATGATGAGTTCATTCAGCGAGTGTG 691  
QY 761 TACGAAACAAAGACTTCCTCCAGACAGCTGAAATCCAGCTCTATGATGCGATGCT 820  
DB 692 TCGGAAACAAAGCTTCCTCCAGACAGCTGAAATCCAGCTCTATGATGCGATGCT 751  
QY 821 TTTATAGATCATGACAGATTAATGATTTATCTGACTTTTGAAGATGAGTGTGATG 880  
DB 752 TTTTCAAGATCTTCGCGCAATATGATTTATCTTGTCTTTAAGATGAGGAGTGTG 811  
QY 881 ACTACATCACTGAGAAATTTGAGAGCCACTGAAACTGAGGAGTGCCTGATATTAG 940  
DB 812 ACTACATCACTGAGAAATTTGAGAGCCACTGAAACTGAGGAGTGCCTGATATTAG 871  
QY 941 GATCCCCACATCACTGAGAAATTTGAGAGCCACTGAAACTGAGGAGTGCCTGAT 1000  
DB 872 GATCCCCACATCACTGAGAAATTTGAGAGCCACTGAAACTGAGGAGTGCCTGAT 931  
QY 1001 TTTCTACCCACAGGAACTGGGAGGTTAATCATGACGATGATGATGATGATGAT 1060  
DB 932 TTTCTACCCACAGGAACTGGGAGGTTAATCATGACGATGATGATGATGATGAT 991  
QY 1061 ATGAGGCTATGTAATGAGAACTGAGAGTGAATCTTAACAGGCACTTTCAGAG 1120  
DB 992 ACAGGCTTACATAGATGAGAACTGAGAGGCAAGTTCGATCAGGCACTTTCAG 1051  
QY 1121 CTCTCAGGGAACGGAATGGGAGTGCAGAGCTCAACAGAGCAATTCATCATGAT 1180  
DB 1052 CGCTCAGGGAACGGAATGGGAGTGCAGAGCTCAACAGAGCAATTCATCATGAT 1111  
QY 1181 TTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1240  
DB 1112 TTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1171  
QY 1241 CACCCAAAATGAGAGGAGAGAAATACCACTGATGATGATGATGATGATGATGAT 1300  
DB 1172 CACCCAAAATGAGAGGAGAGAAATACCACTGATGATGATGATGATGATGATGAT 1231  
QY 1301 CTCTCTACACCTCCGAGCTCCACCTTGTAGCTCTTTGAGAGATGATGATGATGAT 1360  
DB 1232 CTCTCTACACCTCCGAGCTCCACCTTGTAGCTCTTTGAGAGATGATGATGATGAT 1288  
QY 1361 TTGAACATCAACGAAAGAGCCAGAGCACTAAGGTGTGTGATGATGATGATGAT 1420  
DB 1289 TCACGATCACTCAAGAAAGAGCCCGGCACTCAGGTGTGTGTGATGATGATGATGAT 1348

|                            |      |   |   |
|----------------------------|------|---|---|
| Db                         | 1124 | TTTCTCACCACCCAGGAACTGGCAAGTTACATCAGACGACTGATTCGTATGACAGATATGT | 1383  |
| Qy                         | 1061 | ATGAGGCTATGTAGATGGAGCTGAGAGCTGAGATCTCTACACGAGGACTCTGACAG      | 1120  |
| Db                         | 1384 | ATGAGGCTATGTAGATGGAGCTGAGAGCTGAGATCTCTACACGAGGACTCTGACAG      | 1443  |
| Qy                         | 1121 | CTCTCAGGAAACGAAATGGGAGTGCACACGTCACACGACCAATTAATCATGATGCAT     | 1180  |
| Db                         | 1444 | CTCTCAGGAAACGAAATGGGAGTGCACACGTCACACGACCAATTAATCATGATGCAT     | 1503  |
| Qy                         | 1181 | TTGAGTGTATGTGTGTCACCAAGCTGTGGCTTAATTCAGCTTTCAGGAAAAG          | 1234  |
| Db                         | 1504 | TTGAGTGTATGTGTGTCACCAAGCTGTGGCTTAATTCAGCTTTCAGGAAAAG          | 1557  |
| RESULT 8                   |      |   |   |
| HSAS35839                  |      | 1651 bp   | mRNA  |
| LOCUS                      |      |   | linear  |
| DEFINITION                 |      |   | PR1 01-FEB-2003   |
| DESCRIPTION                |      |   | Homo sapiens mRNA for putative alpha3-fucosyltransferase (FUT10 gene), inactive short splice variant. |
| ACCESSION                  |      |   | AU535839  |
| VERSION                    |      |   | AU535839.1 GI:27475658  |
| KEYWORDS                   |      |   | alpha3-fucosyltransferase; alternative splicing; FUT10 gene.  |
| SOURCE                     |      |   | Homo sapiens (human)  |
| ORGANISM                   |      |   | Homo sapiens  |
| REFERENCE                  |      |   | 1   |
| AUTHORS                    |      |   | Candelier, J. J., Martinez-Duncker, I., Oriol, R. and Mollicon, R.                                    |
| TITLE                      |      |   | Cloning expression and genomic organization of two new human  |
| JOURNAL                    |      |   | alpha3-fucosyltransferases (FUT10 and FUT11)  |
| REFERENCE                  |      |   | 2 (bases 1 to 1651)   |
| AUTHORS                    |      |   | Oriol, R.   |
| TITLE                      |      |   | Direct Submission   |
| JOURNAL                    |      |   | Submitted (27-DEC-2002) Oriol R., US04, INSERM, 16 Avenue Paul  |
| COMMENT                    |      |   | Vaillant-Couturier, 94807, Villejuif, FRANCE  |
| FEATURES                   |      |   | related sequences: AU512465, AU535838.  |
| SOURCE                     |      |   | Location/Qualifiers   |
| gene                       |      |   | 1. 1651   |
| CDS                        |      |   | /organism="Homo sapiens"  |
|                            |      |   | /db_xref="taxon:9606"   |
|                            |      |   | /mol_type="mRNA"  |
|                            |      |   | 1. 1651   |
|                            |      |   | /gene="FUT10"   |
|                            |      |   | 559. 1632   |
|                            |      |   | /gene="FUT10"   |
|                            |      |   | /function="unknown"   |
|                            |      |   | /note="inactive short splice variant, lacks the   |
|                            |      |   | transmembrane domain and the first and half of the second   |
|                            |      |   | conserved peptide motifs"   |
|                            |      |   | /codon_start=1  |
|                            |      |   | /product="putative alpha3-fucosyltransferase"   |
|                            |      |   | /protein_id="CAD59772.1"  |
|                            |      |   | /db_xref="GI:27475659"  |
|                            |      |   | /db_xref="GOA:O81V16"   |
|                            |      |   | /db_xref="SPTREMBL:O81V16"  |
|                            |      |   | /translation="MGIGQLPHYALVADGDEVRPMRSLFLHHOPDLPPSSHD  |
|                            |      |   | OSIFILMSQDESPPLAENDTPNIDSLPLPKAHDAVAVHESPKNNYCLFKRPVIT  |
|                            |      |   | TEFNATFSRSHLPULTYLESIEVLSKRYLVLPDSKRLKRLAPLVFQSDSD  |
|                            |      |   | PSDRDSYVRELMTYIEVDYSGECLRNKDLPOQLNPSAMDADGFRILIAQKFLIF  |
|                            |      |   | ENAVCDYITEKWRPLKGVVYVYSGPSITDMLPSKSA1LVSESHPEELASYR   |
|                            |      |   | RLDSDRLYEAVYEMKLKEISNORLLTALBRKRWGVQDVAVIDAFECMVCTKVM   |
|                            |      |   | ANIRLQEKVSKSVGIIEPAGMPS"  |
| ORIGIN                     |      |   |   |
| Query Match                |      |   | 74.2%; Score 1084.4; DB 9; Length 1651;   |
| Best Local Similarity      |      |   | 96.7%; Pired. No. 0;  |
| Matches 1133; Conservative |      |   | 0; Mismatches 1; Indels 38; Gaps 1;   |
| Qy                         | 101  | AGGTCATGTTGAGCTGGGAAAGTTGAAGGAGGTTTAAAGTTCAGTTGCAAG           | 160   |
|                            |      |   |   |
| Db                         | 414  | AGGTCATGTTGAGCTGGGAAAGTTGAAGGAGGTTTAAAGTTCAGTTGCAAG           | 473   |
| Qy                         | 161  | ATGACATACAAAATGAGAGAGACCTAGCATCTTAATTCATTTCTTAAGAAAG          | 220   |
| Db                         | 474  | ATGACATACAAAATGAGAGAGACCTAGCATCTTAATTCATTTCTTAAGAAAG          | 533   |
| Qy                         | 221  | GATTGACCTTCAACGAGAAAGAAATGGAATTGAGAGCTTACCCATTATGCTTGCT       | 280   |
| Db                         | 534  | GATTGACCTTCAACGAGAAAGAAATGGAATTGAGAGCTTACCCATTATGCTTGCT       | 593   |
| Qy                         | 281  | GGTCCCGCTGACGGGGGAGACTGAGAGTTAGCCAAATGAGAGAGATCTGTTCT         | 340   |
| Db                         | 594  | GGTCCCGCTGACGGGGGAGACTGAGAGTTAGCCAAATGAGAGAGATCTGTTCT         | 653   |
| Qy                         | 341  | TCACCATCAACCGGACCTAACCCTCATCATCATGACCAAGCAATTCCTCTTCT         | 394   |
| Db                         | 654  | TCACCATCAACCGGACCTAACCCTCATCATCATGACCAAGCAATTCCTCTTATGCTC     | 713   |
| Qy                         | 395  | -----ATGGTACGACTTTAATAGTAGCTT                                 | 422   |
| Db                         | 714  | CCAAGAGATACCACTTTGGCAGAAAATGATGTAAGTCTTAACATAGTAGCTT          | 773   |
| Qy                         | 423  | ACCTGCTGCTGGAAAGCCATCATGACTGGGCTGTTTTCATGAGAGTCCCGAAA         | 482   |
| Db                         | 774  | ACCTGCTGCTGGAAAGCCATCATGACTGGGCTGTTTTCATGAGAGTCCCGAAA         | 833   |
| Qy                         | 483  | CAATTATAGCTCTTTCATTAACCGATTAACCTTGTTCACATCACTGCGACGTTG        | 542   |
| Db                         | 834  | CAATTATAGCTCTTTCATTAACCGATTAACCTTGTTCACATCACTGCGACGTTG        | 893   |
| Qy                         | 543  | CAGGCAATCCCACTTCCCACTACTACCAATTAACCTTGGAGAGCATTAAGCTGAA       | 602   |
| Db                         | 894  | CAGGCAATCCCACTTCCCACTACTACCAATTAACCTTGGAGAGCATTAAGCTGAA       | 953   |
| Qy                         | 603  | ACTCGATACCTAGTCTCTTTCAGTCCCAAGCAAGCTTGAAGAAAGCTTCCGCT         | 662   |
| Db                         | 954  | ACTCGATACCTAGTCTCTTTCAGTCCCAAGCAAGCTTGAAGAAAGCTTCCGCT         | 1013  |
| Qy                         | 663  | GGGTATGATGACGATGACGATGACGACCAATCAAGGAGAGGAGCTATGTCGAGCT       | 722   |
| Db                         | 1014 | GGGTATGATGACGATGACGATGACGACCAATCAAGGAGAGGAGCTATGTCGAGCT       | 1073  |
| Qy                         | 723  | GATGATTAACATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG       | 782   |
| Db                         | 1074 | GATGATTAACATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG       | 1133  |
| Qy                         | 783  | GCAAGTGAAGAAATCAAGCTCTATGATGATGATGATGATGATGATGATGATGATG       | 842   |
| Db                         | 1134 | GCAAGTGAAGAAATCAAGCTCTATGATGATGATGATGATGATGATGATGATGATG       | 1193  |
| Qy                         | 843  | TAAAGTTATCTAGCTTTTGAAGATGAGCTTGTGATGATGATGATGATGATGATG        | 902   |
| Db                         | 1194 | TAAAGTTATCTAGCTTTTGAAGATGAGCTTGTGATGATGATGATGATGATGATG        | 1253  |
| Qy                         | 903  | GAGGCCACTGAAACCTGGGGGATGCTGCTGATATTAAGGATCCCGACATCAAGCTG      | 962   |
| Db                         | 1254 | GAGGCCACTGAAACCTGGGGGATGCTGCTGATATTAAGGATCCCGACATCAAGCTG      | 1313  |
| Qy                         | 963  | GGTCCCAAGTAAAGAGGATATCTGATGATGATGATGATGATGATGATGATGATG        | 1022  |
| Db                         | 1314 | GGTCCCAAGTAAAGAGGATATCTGATGATGATGATGATGATGATGATGATGATG        | 1373  |
| Qy                         | 1023 | AAGTTACATCAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG        | 1082  |
| Db                         | 1374 | AAGTTACATCAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG        | 1433  |
| Qy                         | 1083 | GCTGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG         | 1142  |
| Db                         | 1434 | GCTGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG         | 1493  |
| Qy                         | 1143 | AGTGAAGAGCTCAACAGAGCAATTAATGATGATGATGATGATGATGATGATGATG       | 1202  |
| Db                         | 1494 | AGTGAAGAGCTCAACAGAGCAATTAATGATGATGATGATGATGATGATGATGATG       | 1553  |

| Db  | 1247 | TGTGATGACTACATACACTGAGAGTTCTGGAGGCCACTGAACTGGAGGGTAACTCCCTGTA  | 1306 |
|---|------|--|------|
| Qy  | 934  | TATTACGGATCCCCCAGATACACAGACTCGGTTCTTCAGAGTAACAAAGTCTATTCTTGTGA | 993  |
| Db  | 1307 | TATTAGGATCCCCCAGATACACAGACTCGGTTCTTCAGAGTAACAAAGTCTATTCTTGTGA  | 1366 |
| Qy  | 994  | TCAGATTTTCTCACCCAGGAACTGGCAAGTTACATCAGACGACTGGATTCTGATAC       | 1053 |
| Db  | 1367 | TCGATATTTTCTCACCCAGGAACTGGCAAGTTACATCAGACGACTGGATTCTGATAC      | 1426 |
| Qy  | 1054 | AGATTGTATGAGGCTTATGTAGATGAGAGCTGAAAGGTGATGATCTTACACGACACTT     | 1113 |
| Db  | 1427 | AGATTGTATGAGGCTTATGTAGATGAGAGCTGAAAGGTGATGATCTTACACGACACTT     | 1486 |
| Qy  | 1114 | CTACACGCTCTCAGGAAACGGAAATGGGAGTGTCAACAGCTCAACACGAGCAATTACATC   | 1173 |
| Db  | 1487 | CTACACGCTCTCAGGAAACGGAAATGGGAGTGTCAACAGCTCAACACGAGCAATTACATC   | 1546 |
| Qy  | 1174 | GATGATTTTGAATGTATGTTGTGACCAAGGTGTGGCTATATACAGGCTTCAGGAAAG      | 1233 |
| Db  | 1547 | GATGATTTTGAATGTATGTTGTGACCAAGGTGTGGCTATATACAGGCTTCAGGAAAG      | 1606 |
| Qy  | 1234 | G 1234   |      |
| Db  | 1607 | G 1607   |      |
| RESULT 7  |      |  |      |
| HSA512465   |      |  |      |
| LOCUS DEFINITION  |      |  |      |
| HSA512465 1623 bp mRNA linear PRI 02-JAN-2003                       |      |  |      |
| Homo sapiens mRNA for putative alpha 1,3-fucosyl transferase (FUT10 |      |  |      |
| gene).  |      |  |      |
| ACCESSION   |      |  |      |
| AJ512465 GI:24370972  |      |  |      |
| VERSION   |      |  |      |
| alpha 1,3-fucosyl transferase; FUT10 gene.                          |      |  |      |
| KEYWORDS  |      |  |      |
| Homo sapiens (human)  |      |  |      |
| SOURCE  |      |  |      |
| Homo sapiens  |      |  |      |
| ORGANISM  |      |  |      |
| Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;   |      |  |      |
| Eukaryota; Euthera; Primates; Catarrhini; Hominiidae; Homo.         |      |  |      |
| REFERENCE   |      |  |      |
| 1 Candelier, J., Martinez-Duncker, I., Oriol, R. and Mollicone, R.  |      |  |      |
| Cloning and genomic organization of a new alpha3-fucosyltransferase |      |  |      |
| FUT10   |      |  |      |
| JOURNAL   |      |  |      |
| Unpublished   |      |  |      |
| REFERENCE   |      |  |      |
| 2 (bases 1 to 1623)   |      |  |      |
| AUTHORS   |      |  |      |
| Oriol, R.   |      |  |      |
| TITLE   |      |  |      |
| Direct Submision  |      |  |      |
| JOURNAL   |      |  |      |
| Submitted (16-OCT-2002) Oriol R., US04, INSEEM, 16 Av. Paul         |      |  |      |
| Vaillant-Conturier, 94897 Villejuif, FRANCE                         |      |  |      |
| COMMENT   |      |  |      |
| related sequences: AJ535839, AJ555839.                              |      |  |      |
| FEATURES  |      |  |      |
| source  |      |  |      |
| 1..1623   |      |  |      |
| /organism="Homo sapiens"  |      |  |      |
| /mol_type="mRNA"  |      |  |      |
| /db_xref="taxon:9606"   |      |  |      |
| /chromosome="8"   |      |  |      |
| /map="8p11.23"  |      |  |      |
| /dev_stage="40 to 70 day embryo"                                    |      |  |      |
| /country="France"   |      |  |      |
| 1..1623   |      |  |      |
| /gene="FUT10"   |      |  |      |
| 345..1604   |      |  |      |
| /gene="FUT10"   |      |  |      |
| /codon_start=1  |      |  |      |
| /evidence=experimental  |      |  |      |
| /product="putative alpha 1,3-fucosyl transferase"                   |      |  |      |
| /protein_id="CAD54669.1"  |      |  |      |
| /db_xref="GI:24370973"  |      |  |      |
| /db_xref="GGA:081VJ3"   |      |  |      |
| /db_xref="SPTREMBL:Q81VJ3"  |      |  |      |
| /translation="MWRIQRRKLASCLVTATVFLVLVLQVNLGKFERKPKSSSS              |      |  |      |
| LDGDKWEEAPTHLNSFLKKEGLTFRKKRWLSDYPTLWWSPLTGFTGLDQCQCGG              |      |  |      |

| Query Match | Best Local Similarity | Matches 1133;   | Score 77.5%; | Conservative 0; | Mismatches 0; | Indels 1; | Gaps 0; | DB 9; | Length 1623; |
|-------------|-----------------------|---|--------------|-----------------|---------------|-----------|---------|-------|--------------|
| QY          | 101                   | AGGTCATGAGTGGAGCTGGGAAAGTTTGAAGAAGAGTTTAAAGTTTCAAGTTTGAAG     | 160          |                 |               |           |         |       |              |
| Db          | 424                   | AGGTCATGAGTGGAGCTGGGAAAGTTTGAAGAAGAGTTTAAAGTTTCAAGTTTGAAG     | 483          |                 |               |           |         |       |              |
| QY          | 161                   | ATGGACATACAAAAATGAGAGAACCACTACGACATCTTAATTCATTCTTAAAGAAAG     | 220          |                 |               |           |         |       |              |
| Db          | 484                   | ATGGACATACAAAAATGAGAGAACCACTACGACATCTTAATTCATTCTTAAAGAAAG     | 543          |                 |               |           |         |       |              |
| QY          | 221                   | GATTGACCTTCAACGAGAAAAAGAAATGGGAATTGGACAGCTAACCCCATTAATCTGT    | 280          |                 |               |           |         |       |              |
| Db          | 544                   | GATTGACCTTCAACGAGAAAAAGAAATGGGAATTGGACAGCTAACCCCATTAATCTGT    | 603          |                 |               |           |         |       |              |
| QY          | 281                   | GGTCCCGCTGACGGGGGAGACTGGGAGGTTTAGGCCAATGTGAGAGAAATGCTTGTCT    | 340          |                 |               |           |         |       |              |
| Db          | 604                   | GGTCCCGCTGACGGGGGAGACTGGGAGGTTTAGGCCAATGTGAGAGAAATGCTTGTCT    | 663          |                 |               |           |         |       |              |
| QY          | 341                   | TGACCATCAACCGGACCTACCTCCATCATACATGACCAAAAGATCCCTCTTAAGTA      | 400          |                 |               |           |         |       |              |
| Db          | 664                   | TGACCATCAACCGGACCTACCTCCATCATACATGACCAAAAGATCCCTCTTAAGTA      | 723          |                 |               |           |         |       |              |
| QY          | 401                   | CTGACTTAAACATAGATAGCTTACTCTGTGCTTGGAAAAGCCCATCATGACTGGCTGTT   | 460          |                 |               |           |         |       |              |
| Db          | 724                   | CTGACTTAAACATAGATAGCTTACTCTGTGCTTGGAAAAGCCCATCATGACTGGCTGTT   | 783          |                 |               |           |         |       |              |
| QY          | 461                   | TTCATGAAGAAGTCCCGGAAAAACAATTAATAGCTCTTCAATAAACAGAGATTACTGTT   | 520          |                 |               |           |         |       |              |
| Db          | 784                   | TTCATGAAGAAGTCCCGGAAAAACAATTAATAGCTCTTCAATAAACAGAGATTACTGTT   | 843          |                 |               |           |         |       |              |
| QY          | 521                   | TCAACTACACTGCGACGTTTCAGCAGGACATTCGCACTTGGCACAATAACCAATCTTG    | 580          |                 |               |           |         |       |              |
| Db          | 844                   | TCAACTACACTGCGACGTTTCAGCAGGACATTCGCACTTGGCACAATAACCAATCTTG    | 903          |                 |               |           |         |       |              |
| QY          | 581                   | AGAGCATTTGAATCCTGTAAGTCACTCCGATTACTAATGTCCTTTGCAAGTCAAAAACAAG | 640          |                 |               |           |         |       |              |
| Db          | 904                   | AGAGCATTTGAATCCTGTAAGTCACTCCGATTACTAATGTCCTTTGCAAGTCAAAAACAAG | 963          |                 |               |           |         |       |              |
| QY          | 641                   | TTAGAAAAAAGATTGCTCCGCTGGTGTATTAAGTCAGACTGTGTGAACCCACATACACA   | 700          |                 |               |           |         |       |              |
| Db          | 964                   | TTAGAAAAAAGATTGCTCCGCTGGTGTATTAAGTCAGACTGTGTGAACCCACATACACA   | 1023         |                 |               |           |         |       |              |
| QY          | 701                   | GGGACAGCTATGTTGCGAGCTGATGACTTAATCGAGGTGATTCCTATGGTGAATGTT     | 760          |                 |               |           |         |       |              |
| Db          | 1024                  | GGGACAGCTATGTTGCGAGCTGATGACTTAATCGAGGTGATTCCTATGGTGAATGTT     | 1083         |                 |               |           |         |       |              |
| QY          | 761                   | TGCGAAAAAAGAACTCCCTCAGAGAGCTGAAAAATCAGACTTATAGATGCGATGGCT     | 820          |                 |               |           |         |       |              |
| Db          | 1084                  | TGCGAAAAAAGAACTCCCTCAGAGAGCTGAAAAATCAGACTTATAGATGCGATGGCT     | 1143         |                 |               |           |         |       |              |
| QY          | 821                   | TTTATAGATCATTTGCAACAGATATAAGTTTATCCTAGCTTTTGAAGTTCAGTTTGATG   | 880          |                 |               |           |         |       |              |
| Db          | 1144                  | TTTATAGATCATTTGCAACAGATATAAGTTTATCCTAGCTTTTGAAGTTCAGTTTGATG   | 1203         |                 |               |           |         |       |              |
| QY          | 881                   | ACTAATCATCTGAGAAAGTTCTGAGAGGCACTGAAATCTGGGGGTAAAGTCCCTGATATTC | 940          |                 |               |           |         |       |              |
| Db          | 1204                  | ACTAATCATCTGAGAAAGTTCTGAGAGGCACTGAAATCTGGGGGTAAAGTCCCTGATATTC | 1263         |                 |               |           |         |       |              |
| QY          | 941                   | GATCCCCCGACATCACAGACTGCGCTTCCAAGTAAACAAAAGTGCATTTCTTATACAGAT  | 1000         |                 |               |           |         |       |              |
| Db          | 1264                  | GATCCCCCGACATCACAGACTGCGCTTCCAAGTAAACAAAAGTGCATTTCTTATACAGAT  | 1323         |                 |               |           |         |       |              |
| QY          | 1001                  | TTTTTCAACCCAGGGAATGGCAATTATCATAGACGACTGATTTGTATGACAGATGTT     | 1060         |                 |               |           |         |       |              |

Db 1024 GATGACATCACTGAGAGGTTCTGGAGGCCACTGAACTGGGGTACTCCCTTATAT 1083  
Qy 937 TACGGATCCCCAGCATCACAGACTGGCTTCCAGTAACAAAGTCTATTCTGTATCA 996  
Db 1084 TACGGATCCCCAGCATCACAGACTGGCTTCCAGTAACAAAGTCTATTCTGTATCA 1143  
Qy 997 GAATTTCTGACCCAGGGAAGTGGCAAGTTACATGACGCTGGAATTCTGATGACGA 1056  
Db 1144 GAATTTCTGACCCAGGGAAGTGGCAAGTTACATGACGCTGGAATTCTGATGACGA 1203  
Qy 1057 TTGTATGAGGCTATGTAGATGAATGAAGCTGAAGGGTGAATCTTAAACGAGCACTTGTG 1116  
Db 1204 TTGTATGAGGCTATGTAGATGAATGAAGCTGAAGGGTGAATCTTAAACGAGCACTTGTG 1263  
Qy 1117 ACAGCTCTCAGGGAAGGGAATGGGAGTGCAGAGCTGAACACGAGCAATTAATCAT 1176  
Db 1264 ACAGCTCTCAGGGAAGGGAATGGGAGTGCAGAGCTGAACACGAGCAATTAATCAT 1323  
Qy 1177 GCATTTGAGTGTATGTGTGACCAAGCTGTGGGCTAATATCAGGCTTCAGGAAAGGAGC 1236  
Db 1324 GCATTTGAGTGTATGTGTGACCAAGCTGTGGGCTAATATCAGGCTTCAGGAAAGGAGC 1383  
Qy 1237 TTACACACCCAAAGATGGGAGGCAAGATACCCACCTGAGTTGCCAGAGCCCAAGTGTG 1296  
Db 1384 TTACACACCCAAAGATGGGAGGCAAGATACCCACCTGAGTTGCCAGAGCCCAAGTGTG 1443  
Qy 1297 TTGGCTTTCTACACATCCCGGAGCTCCACCTTTGAGCTCTTTCGAGAGATGTGATTTCC 1356  
Db 1444 TTGGCTTTCTACACATCCCGGAGCTCCACCTTTGAGCTCTTTCGAGAGATGTGATTTCC 1503  
Qy 1357 AGCTTTGAACATCCAAAGAAAGAGCCCAAGGCACTAAGGTGCTGTTGATAGATCAA 1416  
Db 1504 AGCTTTGAACATCCAAAGAAAGAGCCCAAGGCACTAAGGTGCTGTTGATAGATCAA 1563  
Qy 1417 AACTTTTCACTCAAGAGTTTGGGGCTAGTATTTCAAGAGACTGA 1461  
Db 1564 AACTTTTCACTCAAGAGTTTGGGGCTAGTATTTCAAGAGACTGA 1608

RESULT 6  
HSA353838 1673 bp mRNA linear PRI 01-FEB-2003  
LOCUS  
DEFINITION Homo sapiens mRNA for putative alpha3-fucosyltransferase (FUT10  
gene), splice variant 1.  
ACCESSION AJ535838  
VERSION AJ535838.1 GI:27475656  
KEYWORDS alpha3-fucosyltransferase; alternative splicing; FUT10 gene.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Candelier,J.J., Martinez-Duncker,I., Oriol,R. and Moillone,R.  
TITLE Cloning, expression and genomic organization of two new human  
JOURNAL alpha3-fucosyltransferases (FUT10 and FUT11)  
REFERENCE Unpublished  
AUTHORS Oriol,R.  
TITLE Direct Submission  
JOURNAL Submitted (27-DEC-2002) Oriol R., U504, INSERM, 16 Avenue Paul  
COMMENT Villiant-Couturier, 94807, Villejuif, FRANCE  
FEATURES  
source  
1. 1673  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1. 1673  
/gene="FUT10"  
4/9. 1654  
/gene="FUT10"  
/function="adds fucose via alpha 1-3 linkage to GlcNAc"

/note="splice variant 1, lacks the transmembrane domain"  
/codon\_start=1  
/product="putative alpha3-fucosyltransferase"  
/protein\_id="CAD5977.1"  
/db\_xref="GI:27475657"  
/db\_xref="GOA:O81V17"  
/translation="MVELGKFERKPPSSSLQDHTMEBAPTHLNSFLKKEGLTNR  
RKWELDSVPIMLWMSPLTGETGLGCGADACPTTINRTYLIHHMKAPLPGCTDEN  
IDSLPFRKHHDRVHRESPPKNYKLFKRPVLTENYATRSRSHLPLTGYLES  
LEVLKSLRYLVLPVQSDCDPSPDRUSVYRELNTYLEVSYGSC  
LNNKDLPOOLNPNPAMDADGYRIIAQKFLIARENAVCDYIITEKFWRLKGLVSV  
YVGSPIIDWLPNSKSAIIVSEFSPHRLASVIRLSDSDRLYEAVYEWMLKEGISNQ  
RLTLRLRRKRGVQDVNDNYIDAFECWCTKWMANIRLEKVSKEVLEPAGWP5"

ORIGIN  
Query Match 77.5%; Score 1133; DB 9; Length 1673;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1136; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 94 ATAAAAGAGTCATGCTGAGCTGGGGAAGTTTGAAGAGAGAGTTTAAAGTTCAGT 153  
Db 467 AATAGAGCTATGCTGAGCTGGGGAAGTTTGAAGAGAGAGTTTAAAGTTCAGT 526  
Qy 154 TTGCAAGATGACATACAAAATGGAGGAGCCCTACGATCTTAATTCATTTCTTAAG 213  
Db 527 TTGCAAGATGACATACAAAATGGAGGAGCCCTACGATCTTAATTCATTTCTTAAG 586  
Qy 214 AAAGAAGATTTGACCTTGAACAGAAAGAAATGGGAATGGACACCTACCCATATG 273  
Db 587 AAAGAAGATTTGACCTTGAACAGAAAGAAATGGGAATGGACACCTACCCATATG 646  
Qy 274 CTCTGGTGGTCCCGCTGACGAGGAGAGCTGAGGAGTTAGGCGCATGTGAGCAGATGCT 333  
Db 647 CTCTGGTGGTCCCGCTGACGAGGAGAGCTGAGGAGTTAGGCGCATGTGAGCAGATGCT 706  
Qy 334 TGTTCCTTCACTATCAACCGGACCTTCCATCATCATCATGACACCAAGATTCCTTTC 393  
Db 707 TGTTCCTTCACTATCAACCGGACCTTCCATCATCATCATGACACCAAGATTCCTTTC 766  
Qy 394 TATGGAAGTGAATTAATGATAGTACTTACCTCGCTCGGAGGAGCCATCATGATGG 453  
Db 767 TATGGAAGTGAATTAATGATAGTACTTACCTCGCTCGGAGGAGCCATCATGATGG 826  
Qy 454 GCTGTTTTTATGAGAGTCCCGGAGAAACATTAATAGCTCTTTCATTAACCGATGATT 513  
Db 827 GCTGTTTTTATGAGAGTCCCGGAGAAACATTAATAGCTCTTTCATTAACCGATGATT 886  
Qy 514 ACCTGTTCACTAAGTACGACGAGCTGACGAGGAGATCCCACTGGACATCAACCCAA 573  
Db 887 ACCTGTTCACTAAGTACGACGAGCTGACGAGGAGATCCCACTGGACATCAACCCAA 946  
Qy 574 TACTTGAAGAGATTAAGTCTGAGAGTCACTCCGATACCTAGTTCCTTTCAGTCCAAA 633  
Db 947 TACTTGAAGAGATTAAGTCTGAGAGTCACTCCGATACCTAGTTCCTTTCAGTCCAAA 1006  
Qy 634 AACCAAGCTTAAGAAAGCTTGCTCCGCTGGTGTATGATGACGTAGACCTGACCCCA 693  
Db 1007 AACCAAGCTTAAGAAAGCTTGCTCCGCTGGTGTATGATGACGTAGACCTGACCCCA 1066  
Qy 694 TCAGACAGGAGCAGTATGTTGGGAGCTGATGATGATACATCGAGGTGATTCCTATGCT 753  
Db 1067 TCAGACAGGAGCAGTATGTTGGGAGCTGATGATGATACATCGAGGTGATTCCTATGCT 1126  
Qy 754 GAATGTTTACGAACAAAGACCTCCCTCAGCAGCTGAAGAAATCCAGCTCTATGATGCC 813  
Db 1127 GAATGTTTACGAACAAAGACCTCCCTCAGCAGCTGAAGAAATCCAGCTCTATGATGCC 1186  
Qy 814 GATGCTTTTATAGATATGACAGATTAAGTTATTCCTAGCTTTTGAAGATGACGATT 873  
Db 1187 GATGCTTTTATAGATATGACAGATTAAGTTATTCCTAGCTTTTGAAGATGACGATT 1246  
Qy 874 TGTGATGACTACATCACTGAGAGTCTTGAGAGCCACTGAAGAACTGGGGGTATGTCCTGTA 933

|            |   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|------------|---|--|--------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| Db         | 1336  | TTTCTCACCACCGGAACTGGCAAGTTACATCAGACGACTGGAATCTGTATGACAAATGTT         |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Qy         | 1061  | ATGAGGCGCTATGTAGAAATGAAATGAAAGCTGAAGGCTGAGATCTCTTAAACGAGCACTTCTGACAG |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Db         | 1396  | ATGAGGCGCTATGTAGAAATGAAATGAAAGCTGAAGGCTGAGATCTCTTAAACGAGCACTTCTGACAG |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Qy         | 1121  | CTCTCAGGGGAACGGAATATGGGGAATGTGCAAGACGTCAACCAACGAGCAATTTCATTCAGATGAT  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Db         | 1456  | CTCTCAGGGGAACGGAATATGGGGAATGTGCAAGACGTCAACCAACGAGCAATTTCATTCAGATGAT  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Qy         | 1181  | TTGAGTGAATGGTGTGCAACCAAGGTGTGGGCTAAATATCAGGCTTCAGAAAAAGGCGCTTAC      |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Db         | 1516  | TTGAGTGAATGGTGTGCAACCAAGGTGTGGGCTAAATATCAGGCTTCAGAAAAAGGCGCTTAC      |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Qy         | 1241  | CACCCAAAAGATGGGAGGCGAAGATATCCACCTGAGTTGGCCAGAGCCCAAGTCTTG            |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Db         | 1576  | CACCCAAAAGATGGGAGGCGAAGATATCCACCTGAGTTGGCCAGAGCCCAAGTCTTG            |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Qy         | 1301  | CTTCTCACCACTCCGCACTCCACCTTGAAGCTCTTTCGACAGATGTGATTTCCAGCT            |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Db         | 1636  | CTTCTCACCACTCCGCACTCCACCTTGAAGCTCTTTCGACAGATGTGATTTCCAGCT            |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Qy         | 1361  | TTGACATATCCAAAGAAAGAGCCCAAGGCACTAAGGTGGCTGGTGTATAGAAATCAAACT         |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Db         | 1696  | TTGACATATCCAAAGAAAGAGCCCAAGGCACTAAGGTGGCTGGTGTATAGAAATCAAACT         |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Qy         | 1421  | TTTCATCTCAAGAGTTTGGGGCTAGATTCAGAGACTGA                               | 1461   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Db         | 1756  | TTTCATCTCAAGAGTTTGGGGCTAGATTCAGAGACTGA                               | 1796   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| RESULT 5   |   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| LOCUS      | HS431184  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| DEFINITION | 2675 bp   | MRNA   | linear |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| VERSION    | 1   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| KEYWORDS   | putative alpha 1,3-fucosyl  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| SOURCE     | transferrase (FUT10 gene).  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ORGANISM   | Homo sapiens  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| REFERENCE  | 1   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| AUTHORS    | Roos, C., Kolmer, M., Mattila, P. and Renkonen, R.  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| TITLE      | Composition of Drosophila melanogaster proteome involved in fucosylated glycan metabolism               |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| JOURNAL    | J. Biol. Chem. 277 (5), 3168-3175 (2002)  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| MEDLINE    | 21671316  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| PUBMED     | 11698403  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| REFERENCE  | 2   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| AUTHORS    | Roos, C.  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| TITLE      | Direct Submission   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| JOURNAL    | Submitted (15-FEB-2002) Roos C., Bioinformatics, Medice1 Ltd, Hartmanns, 8, FIN-00290 Helsinki, FINLAND |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| FEATURES   | location/Qualifiers   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| SOURCE     | 1. 2675   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /organism="Homo sapiens"  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /mol_type="mRNA"  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /db_xref="taxon:9606"   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /clone="CSDD001YG24"  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /sex="male"   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /cell_type="neuroblastoma"  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /tissue_type="brain"  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /clone_id="LTI_NFL001_NBC4"   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | 1. 2675   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /gene="FUT10"   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | 1. 321  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /gene="FUT10"   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | 322. 1608   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /gene="FUT10"   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /codon_start=1  |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|            | /product="putative alpha 1,3-fucosyl transferase"   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CDS        |   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5' UTR     |   |  |        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

| Query Match | Best Local Similarity | Score  | DB   | Length |
|-------------|-----------------------|--------|------|--------|
| 97          | 92.8%                 | 1355.4 | DB 9 | 2675   |
| 244         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 157         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 304         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 217         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 364         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 277         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 424         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 337         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 484         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 397         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 544         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 457         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 604         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 517         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 664         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 577         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 724         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 637         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 784         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 697         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 844         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 757         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 904         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 817         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 964         | 99.6%                 | 1355.4 | DB 9 | 2675   |
| 877         | 99.6%                 | 1355.4 | DB 9 | 2675   |

|            |   |   |      |
|------------|---|---|------|
| QY         | 1141  | GAAGTGCAGAGCTCAACCGAGCAATTACATGATGATTTAGTGTGTCAC          | 1200 |
| Db         | 1516  | GAAGTGCAGAGCTCAACCGAGCAATTACATGATGATTTAGTGTGTCAC          | 1575 |
| QY         | 1201  | AAGTGTGGGCTAAATATCAGCTTCAGAAAAGGGCTTACACCCAAAAGATGGAGCA   | 1260 |
| Db         | 1576  | AAGTGTGGGCTAAATATCAGCTTCAGAAAAGGGCTTACACCCAAAAGATGGAGCA   | 1635 |
| QY         | 1261  | GAATATCCCAACCTGATTTGCCAGAGCCCAACAGTGTGTTCTTCCACACTCCGAGCT | 1320 |
| Db         | 1696  | GAATATCCCAACCTGATTTGCCAGAGCCCAACAGTGTGTTCTTCCACACTCCGAGCT | 1695 |
| QY         | 1321  | CCACCTTGAAGCTTTTGCAGAGATGTGATTTCCAGCTTTGAACATCCAAAGAA     | 1380 |
| Db         | 1696  | CCACCTTGAAGCTTTTGCAGAGATGTGATTTCCAGCTTTGAACATCCAAAGAA     | 1755 |
| QY         | 1381  | GCCAGGCACTAAGCTGGCTGTTGATAGATCAAACTTTTCATCTCAAGATTTGG     | 1440 |
| Db         | 1756  | GCCAGGCACTAAGCTGGCTGTTGATAGATCAAACTTTTCATCTCAAGATTTGG     | 1815 |
| QY         | 1441  | GGCTAGTATCAAGACTGA  | 1461 |
| Db         | 1816  | GGCTAGTATCAAGACTGA  | 1836 |
| RESULT 4   |   |   |      |
| LOCUS      | HSAS82015   | 2312 bp   | mRNA |
| DEFINITION | Homo sapiens mRNA for alpha3-fucosyltransferase (FUT10 gene).   |   |      |
| VERSION    | AJ582015  |   |      |
| KEYWORDS   | alpha3-fucosyltransferase; FUT10 gene.  |   |      |
| SOURCE     | Homo sapiens (human)  |   |      |
| ORGANISM   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. |   |      |
| REFERENCE  | Martinez-Duncker, I., Candelier, J. V., Oriol, R. and Mollicone, R.   |   |      |
| AUTHORS    | Cloning, expression and genomic organization of two new human   |   |      |
| JOURNAL    | alpha3-fucosyltransferases (FUT10 and FUT11)  |   |      |
| TITLE      | Unpublished   |   |      |
| REFERENCE  | 2 (bases 1 to 2312)   |   |      |
| AUTHORS    | Oriol, R.   |   |      |
| JOURNAL    | Direct Submission   |   |      |
| TITLE      | Submitted (03-SEP-2003) Oriol R., US04, Inserm, 16 Av. Paul   |   |      |
| COMMENT    | Vailliant-Couturier, 94807, FRANCE  |   |      |
| FEATURES   | Splice variants: AJ512465, AJ535839, AJ535838, AJ431184.  |   |      |
| source     | Location/Qualifiers   |   |      |
|            | 1..2312   |   |      |
|            | /organism="Homo sapiens"  |   |      |
|            | /mol_type="mRNA"  |   |      |
|            | /db_xref="taxon:9606"   |   |      |
|            | /chromosome="8"   |   |      |
|            | /map="8p11.23"  |   |      |
|            | /clone="479"  |   |      |
| gene       | 1..2312   |   |      |
| CDS        | /gene="FUT10"   |   |      |
|            | 357..1796   |   |      |
|            | /functions="adds fucose in alpha3-linkage to GlcNAc"  |   |      |
|            | /codon_start=1  |   |      |
|            | /evidence="experimental"  |   |      |
|            | /product="alpha3-fucosyltransferase"  |   |      |
|            | /protein_id="CAE46499.1"  |   |      |
|            | /db_xref="GI:34481986"  |   |      |
|            | /translation="MVRIOREKLLASCLVATVFLVTLQVWELGFERKEFKSS  |   |      |
|            | LDGHTMEAPTHLNLKKEKGLFPNKRKWELESDYIMMGSLPTEGRGGOCA   |   |      |
|            | DACFTLNRYLHHMTKAPLYGTDENIDSLPLBRKHDMAVHEHESPNNYKLFH   |   |      |
|            | KVITLFWYATFERSHSLPLITTOYLSIEVLKSLRYLVPQSNKTKRKLALVVO  |   |      |
|            | SPDCPSRSDSYRELMTYIEVDSYGECLRNKDPQOLKRPASDADGFRITIAQKF   |   |      |
|            | ILAFENAVDDYITTKFMRPLKGLVPTVGSPTITMLPSNKAIIIVSFSHPRELA   |   |      |
|            | STIRLUDSDRLYEAVVEMKLGKETSQRLTLALERKQGVQVNDQNYIDAECVVC   |   |      |
|            | TKWMANIRLOEKGLPPKWEAEDTHLSCEPTEVAFSPPLRTPLPSLRWMISSFECS   |   |      |
| QY         | 101   | AGTCAATGCTGAGCTGGGAAAGTTGAAGAAGAGATTTAAAGTTCCAGTTGCAAG    | 160  |
| Db         | 436   | AGTCAATGCTGAGCTGGGAAAGTTGAAGAAGAGATTTAAAGTTCCAGTTGCAAG    | 495  |
| QY         | 161   | ATGACATACAAAATGAGAGAACCTACGCACTTATTCATTTCTTAAGAAAG        | 220  |
| Db         | 496   | ATGACATACAAAATGAGAGAACCTACGCACTTATTCATTTCTTAAGAAAG        | 555  |
| QY         | 221   | GATTGACCTTCAACAGAAAAGAAAATGGAAATTGACAGCTACCCATTATGCTTGGT  | 280  |
| Db         | 556   | GATTGACCTTCAACAGAAAAGAAAATGGAAATTGACAGCTACCCATTATGCTTGGT  | 615  |
| QY         | 281   | GGTCCCGCTGACGGGGAGACTGGAGGTTAGCCATGTGAGACAGATCTGTTCT      | 340  |
| Db         | 616   | GGTCCCGCTGACGGGGAGACTGGAGGTTAGCCATGTGAGACAGATCTGTTCT      | 675  |
| QY         | 341   | TACCATCAACCGACACTTCTCCATCATCATGACCAAGATCTCTTCTTAAGTA      | 400  |
| Db         | 676   | TACCATCAACCGACACTTCTCCATCATCATGACCAAGATCTCTTCTTAAGTA      | 735  |
| QY         | 401   | CTGACTTTACATAGTATGCTTACCTCGCTCGGAAAGCCCATCATGACCTGCTTT    | 460  |
| Db         | 736   | CTGACTTTACATAGTATGCTTACCTCGCTCGGAAAGCCCATCATGACCTGCTTT    | 795  |
| QY         | 461   | TTGATGAGAGTCCCGGAAAAAATTAAGCTTTTCATTAACCAAGTATTAACCTGT    | 520  |
| Db         | 796   | TTGATGAGAGTCCCGGAAAAAATTAAGCTTTTCATTAACCAAGTATTAACCTGT    | 855  |
| QY         | 521   | TCACTACACGCGCCAGTTCAGAGGACTTCCCATCTGACCTAACCAACACTGG      | 580  |
| Db         | 856   | TCACTACACGCGCCAGTTCAGAGGACTTCCCATCTGACCTAACCAACACTGG      | 915  |
| QY         | 581   | AGAGCATTAAGTCTCTGAGTCACTCCGATCTAGTCTCTTCCAGTCCAAAAAGC     | 640  |
| Db         | 916   | AGAGCATTAAGTCTCTGAGTCACTCCGATCTAGTCTCTTCCAGTCCAAAAAGC     | 975  |
| QY         | 641   | TTAGAAAAAGACTTCTCCGCTGCTATGACGTGACAGTGAACCCATCAGACA       | 700  |
| Db         | 976   | TTAGAAAAAGACTTCTCCGCTGCTATGACGTGACAGTGAACCCATCAGACA       | 1035 |
| QY         | 701   | GGGACAGTATGTTGGCAGCTGATGATCTTACATGAGAGTGCATTTCTTATGTAATGT | 760  |
| Db         | 1036  | GGGACAGTATGTTGGCAGCTGATGATCTTACATGAGAGTGCATTTCTTATGTAATGT | 1095 |
| QY         | 761   | TACGAAACAAAGCTCTCTCAGAGCTGAAAAATCCAGCTCTATGATCCGATGCT     | 820  |
| Db         | 1096  | TACGAAACAAAGCTCTCTCAGAGCTGAAAAATCCAGCTCTATGATCCGATGCT     | 1155 |
| QY         | 821   | TTTATAGATCATGACAGATATAGTTATTCCTAGTCTTTTGAAGCAAGTTTGGATG   | 880  |
| Db         | 1156  | TTTATAGATCATGACAGATATAGTTATTCCTAGTCTTTTGAAGCAAGTTTGGATG   | 1215 |
| QY         | 881   | ACTACATCACTGAGAAAGTTTGGAGGCACTGAAACTGGGGTATGCTCCGTATATACG | 940  |
| Db         | 1216  | ACTACATCACTGAGAAAGTTTGGAGGCACTGAAACTGGGGTATGCTCCGTATATACG | 1275 |
| QY         | 941   | GATCCCGAGATCAACAGACTGGCTTCCAGTAAACAAATGCTATTTCTTATGCAAT   | 1000 |
| Db         | 1276  | GATCCCGAGATCAACAGACTGGCTTCCAGTAAACAAATGCTATTTCTTATGCAAT   | 1335 |
| QY         | 1001  | TTTCTACCCAGGAACTGGCAAGTTATACAGACAGCTGATTCATGACAGATTGT     | 1060 |



REFERENCE 2 (bases 1 to 3610)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (08-DEC-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nih.gov](mailto:nisc_mgc@nih.gov)  
Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Grant, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, J., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McColley, J.C., McDowell, J., Pearson, R., Stantford, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

FEATURES  
source  
1. 3610  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:74714 IMAGE:5198050"  
/tissue\_type="Brain, adult, 6 pooled whole brains"  
/clone\_id="NTH\_MGC\_114"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
271..1836  
/codon\_start=1  
/product="unknown (protein for MGC:74714)"  
/db\_xref="gi:39645817"  
/translation="MARIQRKLLASGLCTATVILLVQLDIVEINIKMTGPQGVTDSQCNDQPLNSTRSSSEHKEVVELGKEREFSSSLODHTMEAPPHN SFLKEGLTFNRKRMELDSYPIMLWSPITGTRIGQCGACFTINRLHHM TKAPLFYGTDFNDISLPLPKAHDMVAHFEEISKONYLKHFKPVTLFNATSRH SHLPLTTOYLESIEVLKSLRYLVLOSKKRLKRLAPLVYKSDODPSPDRSVREI MTLVSDSYGCELRNKDLPOLKNPASMADGFRITIAKYKELIAFENAVCDIYTK FMRILKIGVPIVYVGSPIIDMLPSKSKALVSEHSEHSLYTRLSDSDRLIYAK VEMLKKEINQRLITVAIEPKMGVDVODNDITDAFECVCTKWANIRLQEKLP KRWAEHTHLSCEPEVFAFSPRTPLPSLSLRMWSISPEOSKEKQALRWLVNDQN FSSQEFMGLVFKD"

misc\_feature  
643..1467  
/note="Glyco transf 10: Region: Glycosyltransferase family 10 (Fucosyltransferase). This family of Fucosyltransferases are the enzymes transferring fucose from GDP-Fucose to GlcNAc in an alpha,3 linkage. This family is know as glycosyltransferase family 10"  
/db\_xref="CD:pfam00852"

Query Match 99.3%; Score 1451.4; DB 9; Length 3610;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1455; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 ATGAAGTCACGGGCGCTCCCGAGGAGTTACAGATCCATGCAATGCTTCAATGATGAG 60  
Db 376 ATGAAGTCACGGGCGCTCCCGAGGAGTTACAGATCCATGCAATGCTTCAATGATGAG 435

QY 61 TGGCTTTATCTAACACCGAGAGCGGAGACATTAAGAAGTCAATGTTAGCTGGGG 120  
Db CGGCTTTATCTAACACCGAGAGCGGAGACATTAAGAAGTCAATGTTAGCTGGGG 495  
QY 121 AAGTTGAAGGAAGAGGTTTAAAGTTCCAGTTTGCAAGTGAACATCAAAAATGGAG 180  
Db AAGTTGAAGGAAGAGGTTTAAAGTTCCAGTTTGCAAGTGAACATCAAAAATGGAG 555  
QY 131 GAAGACCTAGCATCTTATATCTTTCTTAAGAAAGAGATGACCTTCAACAGAAA 240  
Db GAAGACCTAGCATCTTATATCTTTCTTAAGAAAGAGATGACCTTCAACAGAAA 615  
QY 241 AGAAATGGGAATGGACAGTACCCCATATGCTCGGTGGGTCGCCCGTGAAGGGGGAG 300  
Db AGAAATGGGAATGGACAGTACCCCATATGCTCGGTGGGTCGCCCGTGAAGGGGGAG 675  
QY 301 ACTGGAGGTTAGGCGCAATGTGAAGACAGATCTTTTCTTCCACATCAACCGGACCTAC 360  
Db ACTGGAGGTTAGGCGCAATGTGAAGACAGATCTTTTCTTCCACATCAACCGGACCTAC 735  
QY 361 CTCGATATGACATGACCAAGACATTCCTCTTATAGTACATGACCTTAAATATAG 420  
Db CTCGATATGACATGACCAAGACATTCCTCTTATAGTACATGACCTTAAATATAG 795  
QY 421 TTACCTTGCCCTCGAAGGCCATCATGACTGGGCTGTTTTCTATGAAGATCCCGAAA 480  
Db TTACCTTGCCCTCGAAGGCCATCATGACTGGGCTGTTTTCTATGAAGATCCCGAAA 855  
QY 481 AACATTTAAGCTCTTCTTAACCAAGATGATTCCTGTTCACTACACTGCCACGTTC 540  
Db AACATTTAAGCTCTTCTTAACCAAGATGATTCCTGTTCACTACACTGCCACGTTC 915  
QY 541 AGCAGGATTCACCATTCGCATCACTAACCAATACCTTGAGAGCATTAAGCCCTGAAG 600  
Db AGCAGGATTCACCATTCGCATCACTAACCAATACCTTGAGAGCATTAAGCCCTGAAG 975  
QY 601 TCACTCCGATACCTAGTCTCTTTCGACGTCGCAAAAACAAGCTTGAAGAAAGCTTCCGG 660  
Db TCACTCCGATACCTAGTCTCTTTCGACGTCGCAAAAACAAGCTTGAAGAAAGCTTCCGG 1035  
QY 661 CTGCTGATATGACATGACGTCGACGTCGACCAACATCAAGAGGACAGCATTTTGGCGAG 720  
Db CTGCTGATATGACATGACGTCGACGTCGACCAACATCAAGAGGACAGCATTTTGGCGAG 1095  
QY 721 CTGATGACTTACATGAGAGTGCATTCCTATGCTGAATGTTTACGAACAAGACCTCCCT 780  
Db CTGATGACTTACATGAGAGTGCATTCCTATGCTGAATGTTTACGAACAAGACCTCCCT 1155  
QY 781 CAGAGCTGAAGAAATCAGGCTCTATGATGCGATGCGCTTTATATGATCATTTGACAG 840  
Db CAGAGCTGAAGAAATCAGGCTCTATGATGCGATGCGCTTTATATGATCATTTGACAG 1215  
QY 841 TATAGTTTATCTTACCTTTTGAATGATGATGATGATGATGATGATGATGATGATGATG 900  
Db TATAGTTTATCTTACCTTTTGAATGATGATGATGATGATGATGATGATGATGATGATG 1275  
QY 901 TGAAGGCTCACTGAATCTGGGGGATGCTCTGTATATTAACGATCCCGACATCAAGAC 960  
Db TGAAGGCTCACTGAATCTGGGGGATGCTCTGTATATTAACGATCCCGACATCAAGAC 1335  
QY 961 TGGCTTCAAGTACAAAGATGCTATCTTGTATCAGAAATTTTCCACCCAGGGAATCG 1020  
Db TGGCTTCAAGTACAAAGATGCTATCTTGTATCAGAAATTTTCCACCCAGGGAATCG 1395  
QY 1021 GCAAGTTAATCAAGACGATGATTTGATGATGATGATGATGATGATGATGATGATGATG 1080  
Db GCAAGTTAATCAAGACGATGATTTGATGATGATGATGATGATGATGATGATGATGATG 1455  
QY 1081 AAGCTGAAGGATGAGATCTCTAACAAGAGACTTGAAGAGCTTCAAGGAAACGAAATGG 1140  
Db AAGCTGAAGGATGAGATCTCTAACAAGAGACTTGAAGAGCTTCAAGGAAACGAAATGG 1515



JOURNAL Patent: WO 0232962-A 3 25-APR-2002;  
 Millennium Pharmaceuticals, Inc. (US)  
 Location/Qualifiers  
 FEATURES  
 source  
 1..1461  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"

Query Match 100.0%; Score 1461; DB 6; Length 1461;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1461; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAAGTCAGAGGGCCCTCCCGAGGAGTTACAGACCTCCATGCAATGCTTCAATGTCAG 60
DB 1 ATGAAGTCAGAGGGCCCTCCCGAGGAGTTACAGACCTCCATGCAATGCTTCAATGTCAG 60
QY 61 TGGCTTTATCTTAACACGAGAGCAGAGCAATAAAGAGTCAATGTTAGCTGGAG 120
DB 61 TGGCTTTATCTTAACACGAGAGCAGAGCAATAAAGAGTCAATGTTAGCTGGAG 120
QY 121 AAGTTGAAGAAGAGAGTTTAAAGTTTCAAGTTTGAAGATGAGACATACAAAATGAG 180
DB 121 AAGTTGAAGAAGAGAGTTTAAAGTTTCAAGTTTGAAGATGAGACATACAAAATGAG 180
QY 181 GAAGACCTACGACCTTAATTCATTTCTTAAGAAAGAGATTGACCTTCAACAGAGAA 240
DB 181 GAAGACCTACGACCTTAATTCATTTCTTAAGAAAGAGATTGACCTTCAACAGAGAA 240
QY 241 AGAAATGGGAATTTGAGACAGTACCCCATTTATGCTGCTGGTCCCGCTGACGGGGAG 300
DB 241 AGAAATGGGAATTTGAGACAGTACCCCATTTATGCTGCTGGTCCCGCTGACGGGGAG 300
QY 301 ACTGGAGGTTAGGAGCAGATGCTGTTTCTTCAACATCAACGGGAGCTTAC 360
DB 301 ACTGGAGGTTAGGAGCAGATGCTGTTTCTTCAACATCAACGGGAGCTTAC 360
QY 361 CTCCATCATCATGACCAAGCATTCCTCTTCTATGATGATGATGATGATGATGATGATG 420
DB 361 CTCCATCATCATGACCAAGCATTCCTCTTCTATGATGATGATGATGATGATGATGATG 420
QY 421 TTACCTGCTCGGAGAAAGCCCATCATGATGAGGCTGTTTATATAGAGAGTCCCGAAA 480
DB 421 TTACCTGCTCGGAGAAAGCCCATCATGATGAGGCTGTTTATATAGAGAGTCCCGAAA 480
QY 481 AACAAATTAAGGCTCTTTCATTAACAGAGTATACCTTGTTCACATCACTGCGACGTT 540
DB 481 AACAAATTAAGGCTCTTTCATTAACAGAGTATACCTTGTTCACATCACTGCGACGTT 540
QY 541 AGCAGGCTTCCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
DB 541 AGCAGGCTTCCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
QY 601 TCACCTCCGATACCTAGTTCCTTTCGATGCTCAAAAACAGAGTTGAAAAGAGCTTCCG 660
DB 601 TCACCTCCGATACCTAGTTCCTTTCGATGCTCAAAAACAGAGTTGAAAAGAGCTTCCG 660
QY 661 CTGCTGATGATGACGTCAGCTGACCTGACCCACATCAGACAGGAGCAGCTATGTTGG 720
DB 661 CTGCTGATGATGACGTCAGCTGACCTGACCCACATCAGACAGGAGCAGCTATGTTGG 720
QY 721 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
DB 721 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
QY 781 CAGAGCTGAAAAATCCAGCTCTATGATGATGATGATGATGATGATGATGATGATGATG 840
DB 781 CAGAGCTGAAAAATCCAGCTCTATGATGATGATGATGATGATGATGATGATGATGATG 840
QY 841 TATAGTTTATCCAGCTTTTGAAGATGAGTTTGAAGATGAGTTTGAAGATGAGTTTGA 900
DB 841 TATAGTTTATCCAGCTTTTGAAGATGAGTTTGAAGATGAGTTTGAAGATGAGTTTGA 900

```

```

QY 901 TGAAGCCACTGAAACTGGGGGATGTCCTGTTATTTACGATATCCCGACATCACAGAC 960
DB 901 TGAAGCCACTGAAACTGGGGGATGTCCTGTTATTTACGATATCCCGACATCACAGAC 960
QY 961 TGGCTTCCAGTACAAAGAGTGTATTTCTTGTATGAGATTTTCTCAACCCGAGGAACTG 1020
DB 961 TGGCTTCCAGTACAAAGAGTGTATTTCTTGTATGAGATTTTCTCAACCCGAGGAACTG 1020
QY 1021 GCAAGTTATCATCAGACGACTGATTTCTGATGACAGATTTGATGAGGCTTATGAAATG 1080
DB 1021 GCAAGTTATCATCAGACGACTGATTTCTGATGACAGATTTGATGAGGCTTATGAAATG 1080
QY 1081 AAGCTGAAGGCTGAGATCTCTAACGAGGACTTTCAGAGCTCTCAGAGGAGACGAAATG 1140
DB 1081 AAGCTGAAGGCTGAGATCTCTAACGAGGACTTTCAGAGCTCTCAGAGGAGACGAAATG 1140
QY 1141 GGAATGCAAGCTCAACAGAGCAATTAATCATGATGATGATTTGATGATGATGATGATG 1200
DB 1141 GGAATGCAAGCTCAACAGAGCAATTAATCATGATGATGATTTGATGATGATGATGATG 1200
QY 1201 AAGTGTGGGCTTAATATCAGGCTTCAGAGAAAGGCTTACACCCAAAGATGGAAGCA 1260
DB 1201 AAGTGTGGGCTTAATATCAGGCTTCAGAGAAAGGCTTACACCCAAAGATGGAAGCA 1260
QY 1261 GAAGATACCCAGCTGATGCTCCAGAGCCACAGTGTGCTTCTCAACTCCGAGACT 1320
DB 1261 GAAGATACCCAGCTGATGCTCCAGAGCCACAGTGTGCTTCTCAACTCCGAGACT 1320
QY 1321 CCACCTTTGAGCTCTTTGCGAGAGATGATGATTTCCAGCTTTGAAACATCCAGAAAGAA 1380
DB 1321 CCACCTTTGAGCTCTTTGCGAGAGATGATGATTTCCAGCTTTGAAACATCCAGAAAGAA 1380
QY 1381 GCCCAGGACTAAGAGTGGCTGTTGATGAGAAATCAATTTTCACTCAAGAGTTTGG 1440
DB 1381 GCCCAGGACTAAGAGTGGCTGTTGATGAGAAATCAATTTTCACTCAAGAGTTTGG 1440
QY 1441 GGCTAGTATTCAGAGACTGA 1461
DB 1441 GGCTAGTATTCAGAGACTGA 1461

```

RESULT 2  
 AXS92444 1669 bp DNA linear PAT 27-JAN-2003  
 LOCUS AXS92444  
 DEFINITION Sequence 1 from Patent WO0232962.  
 ACCESSION AXS92444  
 VERSION AXS92444.1 GI:27950546  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 source  
 1..1669  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 163..1623  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAD62017.1"  
 /db\_xref="GI:27950547"  
 /db\_xref="RMBTREM1:CAD62017"  
 /translation="MKVTCGPGGYDMSOCFNDQWPLSNTSSSHIKXWVLEKFER  
 KFKSSLDQGHKKBEAPHLNLSFLKKEGLTFNRKRMBLDSYPIKIMWSFLTSETG  
 RLGQCCADACPFITNRTYLNHMTKAPLPYGLDFNIDSLPLPRKXHHWAVAFHESPK  
 NNKYLPHKXEVITLFFNYATPFSRSHSLPTTQYLEISIEVLKSLRYLVPQSKKXKRL"

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: September 12, 2004, 19:48:00 / Search time 3891.02 Seconds

(without alignments)  
16274.436 Million cell updates/sec

Title: US-10-080-960-3

Perfect score: 1461

Sequence: 1 atgaagtcacggcgccctcc.....gacctatccaagactga 1461

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: 1: gb\_da: 2: gb\_htg: 3: gb\_in: 4: gb\_om: 5: gb\_ov: 6: gb\_pat: 7: gb\_ph: 8: gb\_pl: 9: gb\_pr: 10: gb\_ro: 11: gb\_sts: 12: gb\_sy: 13: gb\_un: 14: gb\_vl: 15: em\_da: 16: em\_fun: 17: em\_hum: 18: em\_in: 19: em\_mu: 20: em\_om: 21: em\_or: 22: em\_ov: 23: em\_ph: 24: em\_pat: 25: em\_pl: 26: em\_ro: 27: em\_sts: 28: em\_un: 29: em\_vl: 30: em\_htg\_hum: 31: em\_htg\_inv: 32: em\_htg\_oher: 33: em\_htg\_mus: 34: em\_htg\_pin: 35: em\_htg\_rod: 36: em\_htg\_mam: 37: em\_htg\_vit: 38: em\_sy: 39: em\_htgo\_hum: 40: em\_htgo\_mus: 41: em\_htgo\_oher:

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Match | Query Length | DB ID | Description          |
|------------|--------|-------|--------------|-------|----------------------|
| 1          | 1461   | 100.0 | 1461         | 6     | AX592446 Sequence    |
| 2          | 1461   | 100.0 | 1669         | 6     | AX592444 Sequence    |
| 3          | 1451.4 | 99.3  | 3610         | 9     | BC063462 Homo sapi   |
| 4          | 1359.4 | 93.0  | 2312         | 9     | HSAS582015 Homo sapi |
| 5          | 1355.4 | 92.8  | 2675         | 9     | HSAS411184 Homo sapi |
| 6          | 1133.4 | 77.5  | 1673         | 9     | HSAS535838 Homo sapi |
| 7          | 1132.4 | 74.2  | 1623         | 9     | HSAS12465 Homo sapi  |
| 8          | 1084.4 | 74.2  | 1651         | 9     | HSAS53839 Homo sapi  |
| 9          | 1013.8 | 69.4  | 1389         | 4     | AY344579 Bos tauru   |
| 10         | 838.8  | 57.4  | 104120       | 9     | AC067838 Homo sapi   |
| 11         | 838.8  | 57.4  | 185872       | 9     | AC091144 Homo sapi   |
| 12         | 799.6  | 54.7  | 2874         | 10    | BC030474 Mus muscu   |
| 13         | 799.6  | 54.7  | 2978         | 10    | BC062113 Mus muscu   |
| 14         | 657.4  | 45.0  | 1657         | 5     | CGAS535692 Gallus ga |
| 15         | 626    | 42.8  | 191072       | 2     | AC115775 Mus muscu   |
| 16         | 581.2  | 39.8  | 597          | 6     | AX314120 Sequence    |
| 17         | 507.6  | 34.7  | 1333         | 10    | BC022579 Mus muscu   |
| 18         | 374    | 25.6  | 185812       | 2     | BX640592 Dario rer   |
| 19         | 370.8  | 25.4  | 165781       | 2     | BX321906 Dario rer   |
| 20         | 300    | 20.5  | 1114         | 9     | BC004884 Homo sapi   |
| 21         | 300    | 20.5  | 188980       | 2     | AC090087 Homo sapi   |
| 22         | 298    | 19.7  | 2180         | 9     | AX817125 Sequence    |
| 23         | 288    | 19.3  | 2027         | 9     | AK098505 Homo sapi   |
| 24         | 282    | 19.3  | 2027         | 9     | AK056199 Homo sapi   |
| 25         | 235.6  | 16.1  | 1506         | 5     | AK060609 Fugu rubr   |
| 26         | 230.2  | 15.8  | 450          | 6     | AR420976 Sequence    |
| 27         | 230.2  | 15.8  | 450          | 6     | BD116529 EST and e   |
| 28         | 212.4  | 14.5  | 1866         | 5     | CH153752 Gallus ga   |
| 29         | 198.2  | 13.6  | 1542         | 10    | RN0535753 Rattus no  |
| 30         | 184    | 12.6  | 1714         | 3     | AK114811 Ciona int   |
| 31         | 183.8  | 12.6  | 3211         | 10    | MMU52293 Mus muscu   |
| 32         | 180.8  | 12.4  | 1479         | 6     | AX356085 Sequence    |
| 33         | 180.8  | 12.4  | 2069         | 9     | BC036037 Homo sapi   |
| 34         | 180.8  | 12.4  | 2557         | 6     | AX356083 Sequence    |
| 35         | 179    | 12.3  | 194879       | 2     | AC109163 Mus muscu   |
| 36         | 178.4  | 12.2  | 1335         | 3     | DME3302046 Drosophi  |
| 37         | 176.8  | 12.1  | 1335         | 3     | AY061932 Drosophi    |
| 38         | 166    | 11.4  | 2408         | 6     | AX833606 Sequence    |
| 39         | 166    | 11.4  | 2408         | 6     | AK095482 Homo sapi   |
| 40         | 129.4  | 8.9   | 181589       | 5     | AC146544 Gasterost   |
| 41         | 129.4  | 8.9   | 197879       | 5     | AC146543 Gasterost   |
| 42         | 116.2  | 8.0   | 218747       | 10    | AC121599 Mus muscu   |
| 43         | 109.8  | 7.5   | 249360       | 2     | AC127920 Rattus no   |
| 44         | 109.8  | 7.5   | 250492       | 2     | AC137179 Rattus no   |
| 45         | 109.8  | 7.5   | 265382       | 2     | AC114439 Rattus no   |

## ALIGNMENTS

RESULT 1  
AX592446  
LOCUS AX592446  
DEFINITION Sequence 3 from Patent WO0232962.  
ACCESSION AX592446  
VERSION AX592446.1 GI:27950548  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Glucksmann, M.A. and Meyers, R.  
TITILE 80090, 52874, 52880, 63497, and 33425 methods and compositions of  
human proteins and uses thereof

**This Page Blank (uspto)**



```

D5      408 CRIC04V-----CRISDQPK 422

RESULT 12
UC4591
alpha-1.3 fucosyltransferase (EC 2.4.1.-) - mouse
C:Species: Mus musculus (house mouse)
C:Date: 16-Apr-1996 #sequence_revision 24-May-1996 #text_change 24-Nov-1999
C:Accession: J04591
R:Ozawa, K.; Muramatsu, T.
J. Biochem. 119, 302-308, 1996
A:Title: Molecular cloning and expression of a mouse alpha-1.3 fucosyltransferase gene
/:Reference number: J04591, MUID:97037075, PMID:8882722

```

```

A/Title: sequences and analysis of the human APOB gene
A/Reference number: A86141; MIMD:21016715; EMID:1130712
A/Accession: F96533
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-513 <SNO>
A/Cross-references: GB:AE05173; NID:G10120428; PINN:AA313053.1; GSPDB:GN00141
A/Genetics:
A/Gene: F14022.8
A/Map position: 1

Query Match      7.7%; Score 203; DB 2; Length 513;
Best Local Similarity 27.88%; Pred. No. 7.1e-08; Tadj 30; Cons 51

```

```

Query Match      7.7%;   Score 203;   DB 2;   Length 513;
Best Local Similarity 27.8%;   Pred. No. 7.1e-08;
Matches . 58; Conservative 37; Mismatches 84; Indels 30; Gaps 6;

```

Db 275 SHKFWLAIENTAVESVTEKLFYALDSSGVPVIGASNVQFPPH-SVIDSGKFSMG 333

Qy 339 ELASYIRRLDSDRLYEAYVEWKKEISNQRLTLALRRKMGVQDVNQYDAFECMV 398

Db 334 ELAAYVRLDDEVAAYSEVHAMRCGIMGNVYKTRAVS-----LDTLPCL 379

Qy 399 CTKY 402

Db 380 CEEI 383

## RESULT 9

A54057  
alpha(1,3)-fucosyltransferase (EC 2.4.1.-) 7 precursor - human  
N:Alternate names: leukocyte fucosyltransferase FucT11  
C:Species: Homo sapiens (man)  
C>Date: 02-Aug-1994 #sequence\_revision 02-Aug-1994 #text\_change 20-Apr-2000  
C/Accession: A54057; A53713  
R:Sasaki, K.; Kurata, K.; Funayama, K.; Nagata, M.; Watanabe, E.; Ohta, S.; Hanai, N.; J. Biol. Chem. 269, 14730-14737, 1994  
A>Title: Expression cloning of a novel alpha1,3-fucosyltransferase that is involved in h  
A/Reference number: A54057; MUID:94237894; PMID:8182079  
A/Accession: A54057  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-342 <SAS>  
A/Cross-references: GB:X78031; NID:G516292; PIDN:CAA54962.1; PID:G516293  
R:Natsuka, S.; Gersten, K.M.; Zenita, K.; Kannagi, R.; Lowe, J.B. J. Biol. Chem. 269, 16789-16794, 1994  
A>Title: Molecular cloning of a cDNA encoding a novel human leukocyte alpha-1,3-fucosylt  
A/Reference number: A53713; MUID:94266898; PMID:8207002  
A/Accession: A53713  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-160, 'X', 163-303, 'SY', 306-342 <NAT>  
A/Cross-references: GB:008112; NID:9520463; PIDN:AAA56869.1; PID:G520464  
C/Genetics:  
A/Gene: GDB:FUT7  
A/Cross-references: GDB:373982  
A/Map position: 9pter-9qter  
C/Superfamily: galactoside 3(4)-L-fucosyltransferase  
C/Keywords: glycoprotein; glycosyltransferase; hexosyltransferase  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:33-342/Product: alpha(1,3)-fucosyltransferase 7 #status predicted <MAT>  
F:81,291/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.5%; Score 224; DB 2; Length 342;  
Best Local Similarity 26.2%; Pred. No. 9.5e-10;  
Matches 85; Conservative 52; Mismatches 133; Indels 54; Gaps 13;

Qy 83 WEIDSYR-----IMLWMSPLTGETGRL-----GQCGADACFFINTTYLHHNMT 126

Db 31 WLLGSAPRGTPAPQPTITLIVHWPFTDOPRLPBDCTCRGIGARCHLSANRSL--ASA 88

Qy 127 KALEFGTDPNIDSLP---RKAHDMAVFHEESPKNKYLFHKPVTLTFLNYATSRH 183

Db 89 DAVAFHRELQTRSHLPLAORPRQOPVWAMSMESPSTHGLSH--LRGIFNWALSYRDL 146

Qy 184 SHLPLTQYLESIEVLKSLRYLVLQSKNKLAKRLAPLVVVYQSDCPDPSDRDSVYRELMT 243

Db 147 SDIFVPYGRLE-----PHWGSPRLPAKSRV-----AAWVSNQERQRLARLYQLAP 195

Qy 244 YIEVDSYSECLRNKULPQGLKRPASMDAGFRITIAQYKFLAENAVCDYITEKFWR- 302

Db 196 HLRVDVFGRA-----NGRPLCASCLVPTVAQRYFLSPENSQHRDYITEKFWRN 244

Qy 303 PLKLGAVVPVYGSPTT--DWLPSNKSAILVSEFHPRELASYIRRLDSDRLYEAVEM 360

Db 245 ALVAGVVPVVLGPRATYFAFPAD-AFVHVDLFGSARELAFLTGM--NSRRQGRFFAW 301

Qy 361 KUKGEISNORLLTLALRRKMGVQD 384

Db 302 RDRLRV---RLFTDWERECALCD 322

## RESULT 10

B36340  
alpha(1,3)-fucosyltransferase (EC 2.4.1.-) 4 precursor [validated] - human  
N:Alternate names: CD15; ELAM-1 ligand fucosyltransferase (ELFT); FCT3A; FUC-TIV; mye  
C:Species: Homo sapiens (man)  
C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C/Accession: B36340; A36340; A40976; A41202  
R:Geelz, S.E.; Hession, C.; Goff, D.; Griffiths, B.; Tizard, R.; Newman, B.; Chi-Rose  
Cell 63, 1349-1356, 1990  
A>Title: ELFT: a gene that directs the expression of an ELAM-1 ligand.  
A/Reference number: A36340; MUID:91084863; PMID:1702034  
A/Accession: B36340  
A/Molecule type: mRNA  
A/Residues: 1-405 <GOE1>  
A/Cross-references: GB:M58596; NID:G182068; PIDN:AAA63172.1; PID:G182069  
A/Accession: A36340  
A/Molecule type: mRNA  
A/Residues: 1-86, 'P', 88-405 <LOW>  
A/Cross-references: GB:M65030; NID:G182791; PIDN:AAA92977.1; PID:G1236720  
R:Kumar, R.; Potvin, B.; Muller, W.A.; Stanley, P. J. Biol. Chem. 266, 21777-21783, 1991  
A>Title: Cloning of a human alpha(1,3)-fucosyltransferase gene that encodes ELFT but  
A/Reference number: A41202; MUID:92042084; PMID:1716983  
A/Accession: A40976  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-86, 'P', 88-405 <LOW>  
A/Cross-references: GB:M65030; NID:G182791; PIDN:AAA92977.1; PID:G1236720  
R:Kumar, R.; Potvin, B.; Muller, W.A.; Stanley, P. J. Biol. Chem. 266, 21777-21783, 1991  
A>Title: Cloning of a human alpha(1,3)-fucosyltransferase gene that encodes ELFT but  
A/Reference number: A41202; MUID:92042084; PMID:1716983  
A/Accession: A41202  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-240, 'D', 242-400 <KUM>  
A/Cross-references: GB:S65161; NID:G239005; PIDN:AAB20349.1; PID:G239006  
C/Genetics:  
A/Gene: GDB:FUT4; CD15; FCT3A; FUC-TIV  
A/Cross-references: GDB:131863; OMIM:104220  
A/Map position: 11q21-11q21  
C/Superfamily: galactoside 3(4)-L-fucosyltransferase  
C/Keywords: glycoprotein; glycosyltransferase; hexosyltransferase  
F:1-48/Domain: signal sequence #status predicted <SIG>  
F:49-405/Product: alpha(1,3)-fucosyltransferase 4 #status predicted <MAT>  
F:1,190/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.4%; Score 221; DB 2; Length 405;  
Best Local Similarity 23.8%; Pred. No. 2.1e-09;  
Matches 92; Conservative 39; Mismatches 139; Indels 116; Gaps 16;

Qy 90 IMLWMSPLTGETG-----ELGCGADACFFINTTYLHHNMTKAFLLGYGDF----- 136

Db 66 VILWMEPPGGDSAPRPPDCL--RFINISGRLLITDRA--SYGBAQVLFFHRDLVYKGP 122

Qy 137 -----NIDSLP-----PRKAHDMAVFHEESPKNKYLF 166

Db 123 DMPPEWGIQAHTAEBVDRLVDYEAAAAAALATSSFPBPGQRMWMMNFSPS----- 176

Qy 167 HKPVV-----TLFNYTATFSRHSULPTQYLESIEVLKSLRYLVLQSKNKLAKRLAP 220

Db 177 HSPGLRSIASNLFMWLTLSYRADS-----DVFVGYGLYPPSHHGDPGSLAP 224

Qy 221 -----LVVYQSDCPDPSDRDSVYRELMTYIEVDSYSECLRNKULPQGLKRPASMDADG 273

Db 225 LSRKGLVAMVVISMDERQARVRYHQLSGHYTVDVGRGPGQVPDE-----IG 274

Qy 274 FYRITAOYKFLAENAVCDYITEKFWR-PLKLGAVVPVYG--SPSITDMLPSNKSAIL 330

Db 275 LHTVARYKFLAENSCHLDYITEKLMRNVALAGAVPVVLGPDRANTYERFVRG-AFTH 333



QY 138 IDSLPRKXAHNDMAVFEHESPKNNYKLFHKFVITLFFNYTATFSRH-----SHLPTTQY 192  
DB 134 LP--PPTPPOGQKRWIFWSPNSN-----CHLEALDGYFNLTMYSY 172  
QY 193 LESIEVLKSLRYLV-----LQSKNKLKRLAPLVYQSDCCPPSDRDSYRE 241  
DB 173 RSDSDIFTPYGMLEPWSGQPAHPPLNLAKTEL-----VAMAVSNMKPDSARVRYQSL 226  
QY 242 MTYIEVDSYGECLRNKDLPOQLKNPASMADGFFRILIAQYKFLIAFENAVCDDYTEKEF 301  
DB 227 CAHLKLVYGR--SHKPLPK-----GTMMETLSRKXYLAFFENSLHPDYITTEKLM 274  
QY 302 R-PLKLGVPVYVYGSPTIDM---LPSNKAALVSESHPRELASIYRLDSDRLYEAY 357  
DB 275 RNALAEAAVAVPVLG-PERSNYERFLPPD-AFIHVDQFSKDLARYQLDLDKHARYLSY 332  
QY 358 VEWLTKGEISNQRLITLALRER--KMGVDVYNQDNYIDAFECMCTKYWANITLQ 409  
DB 333 FRWR-----ETLPRPSFMSAL-----AF-CAACWKLQGSRYQ 364

RESULT 7  
A36669  
3-alpha-galactosyl-N-acetylglucosaminide 4-alpha-L-fucosyltransferase (EC 2.4.1.65) - hu  
N/Alternate names: alpha (1,3/1,4) fucosyltransferase; blood group Lewis alpha-4-fucosyl  
C/Species: Homo sapiens (man)  
C/Date: 12-Apr-1991 #sequence\_revision 12-Apr-1991 #text\_change 03-Jun-2002  
C/Accession: A36669; J39043; J39044; J39045; S12123  
R/Kukowska-Latallo, J.F.; Larsen, R.D.; Natf, R.P.; Lowe, J.B.  
Gene Dev 4, 1288-1303, 1990  
A/Title: A cloned human cDNA determines expression of a mouse stage-specific embryonic a  
A/Reference number: A36669; MUID:91032981; PMID:1977660  
A/Accession: A36669  
A/Status: Preliminary  
A/Molecule type: mRNA  
A/Residues: 1-361 <RES>  
A/Cross-references: GB:X53578; NID:g28523; PIDN:CA37641.1; PID:g28530  
R/Cameron, H.S.; Szczepanski, D.; Westcom, B.W.  
J. Biol. Chem. 270, 20112-20122, 1995  
A/Title: Expression of human chromosome 19p alpha(1,3)-fucosyltransferase genes in norma  
A/Reference number: J39043; MUID:95378269; PMID:7650030  
A/Accession: J39043  
A/Status: Preliminary  
A/Molecule type: mRNA  
A/Residues: 1-361 <RES>  
A/Cross-references: EMBL:U27327; NID:g967190; PIDN:AAC50186.1; PID:g967191  
A/Accession: J39045  
A/Status: Preliminary  
A/Molecule type: mRNA  
A/Residues: 1-361 <RES>  
A/Cross-references: EMBL:U27328; NID:g967192; PIDN:AAC50187.1; PID:g967193  
C/Genetics:  
A/Gene: GDB:FUT3; LE  
A/Cross-references: GDB:135717; OMIM:111100  
A/Map position: 19p13.3-19p13.3  
A/Note: alternative splicing 5' to the coding region  
C/Superfamily: galactoside 3(4)-L-fucosyltransferase  
C/Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 9.1%; Score 239.5; DB 2; Length 361;  
Best Local Similarity 26.5%; Pred. No. 6.4e-11;  
Matches 94; Conservative 43; Mismatches 115; Indels 103; Gaps 18;

QY 90 IMLWMSPLTGTGRLGOC-----GADACFTTNR-----TYLHHMTKAPLVYGTDEN 137  
DB 65 ILLMTWPHIIVA-LSRSEWVPSTADCHITARXKVPQADYIVNH-----MD 112

QY 138 IDS-----LPLPRKXAHNDMAVFEHESPKNNYKLFHKFVITLFFNYTATFSRHSHLPLTTQ 191  
DB 113 IMNPKSRLLPSPRPOGQKRWIFWSPNSNCHL--EALDGYFN-----LTMS 158  
QY 192 YLESTIEVLKSLRYLV-----LQSKNKLKRLAPLVYQSDCCPPSDRDSYRE 240  
DB 159 YRSDSDIFTPYGMLEPWSGQPAHPPLNLAKTEL-----VAMAVSNMKPDSARVRYQSL 212  
QY 241 LMTYIEVDSYGECLRNKDLPOQLKNPASMADGFFRILIAQYKFLIAFENAVCDDYTEKEF 300  
DB 213 LQHLKLVYGR--SHKPLPK-----GTMMETLSRKXYLAFFENSLHPDYITTEKLM 260  
QY 301 WR-PLKLGVPVYVYGSPTIDM---LPSNKAALVSESHPRELASIYRLDSDRLYEAY 356  
DB 261 WRNALAEAAVAVPVLG-PERSNYERFLPPD-AFIHVDQFSKDLARYQLDLDKHARYLSY 318  
QY 357 YVEWTKLGEISNQRLITLALRER--KMGVDVYNQDNYIDAFECMCTKYWANITLQ 409  
DB 319 YFRWR-----ETLPRPSFMSAL-----FCKACWKLQGSRYQ 351

RESULT 8  
H96742  
hypotheical protein PTM19.14 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C/Accession: H96742  
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, C.W.; Chung, M.X.; Comai, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.,  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: H96742  
A/Status: Preliminary  
A/Molecule type: DNA  
A/Residues: 1-393 <GTO>  
A/Cross-references: GB:AE005173; NID:g6978923; PIDN:AAF34315.1; GSPDB:GN00141  
C/Genetics:  
A/Gene: PTM19.14  
A/Map position: 1

Query Match 8.8%; Score 230; DB 2; Length 393;  
Best Local Similarity 24.3%; Pred. No. 3.9e-10;  
Matches 103; Conservative 60; Mismatches 159; Indels 102; Gaps 19;

QY 13 SMOCPDQWPLSTRTSESHKXVVELGKE-----RKZFK-----SSLOD-- 54  
DB 28 SMHSIDPRKXLSDSIDSPSDVLAKKWDPEVGCARREHKKALIGNVSSGSLQFEG 67  
QY 55 -GHTKCEADPTLNSFLKKEGLTFNRKRWELSDSYPIKLMWSPLTGFTGRLGCGADACF 113  
DB 68 CGKLKKK--HVKVLVK--GWT-----WIPDMLLEMLYSGRCGM-TCL 123  
QY 114 FTNRTYLHHMTKAPLVYGTDENISLPLPRKXAHNDMAVFEHESPKNNYKLFHKF 169  
DB 124 WT--KSSVLADSPDALFETT-----TPLOQRVQDPLRAVYELGAKKRGREIF-- 173  
QY 170 VITLFFNYTATFSRHSHLPLTTQYLESIEVLKSLRYLVPLQSKNKLKRLAPLVYQSD-C 228  
DB 174 -----ISYHAADVQTYAGSL-PHNNRNYHISPHKNDV-----LVWSSSRC 216  
QY 229 DPSPDRSYRELMYTYIEVDSYGECLRN-KDLPOQL-----NPSAMDADGFYHII 278  
DB 217 LP--HRRLAKSLDIDIPHISFGKLNNGGDSALSWPECVAAHNAFAKRYDHLQCM 274  
QY 279 AOVYFIAPFNAVCDDYTEKEFWRLKLVGVVYVYGSPTIDMLPSNKAALVSEFSHR 338



**!Note:** the authors translated the codon GAC for residue 219 as Asn  
R.Cameron, H.S., Szecepaniak, D., Weston, B.W.  
J. Biol. Chem. 270, 20112-20122, 1995  
A:Title: Expression of human chromosome 19p alpha(1,3)-fucosyltransferase genes in nem  
A:Reference number: 139043; MUID:95378269; PMID:7650030  
A:Accession: 139050  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1359 <RES>  
A:Cross-references: EMBL:U27333; NID:g967202; PIDD:AA050192.1; PID:g967203  
A:Accession: 139051  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1359 <RE2>  
A:Cross-references: EMBL:U27334; NID:g967204; PIDD:AA050193.1; PID:g967205  
A:Accession: 139052  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1359 <RE4>  
A:Cross-references: EMBL:U27336; NID:g967208; PIDD:AA050195.1; PID:g967209  
A:Accession: 139054  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1359 <RE5>  
A:Cross-references: EMBL:U27337; NID:g967210; PIDD:AA050196.1; PID:g967211  
C:Genetics:  
A:Gene: GDB:FUT6  
A:Cross-references: GDB:135180; OMIM:136836  
A:Map position: 19p13.3-19p13.3  
A:Superfamily: galactoside 3(4)-L-fucosyltransferase  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 10.0%; Score 263, DB 2, Length 359;  
Best Local Similarity 27.4%; Pred. No. 9, 5e-13;  
Matches 98; Conservative 48; Mismatches 118; Indels 94; Gaps 19;

QY 87 SYPIIMVWSPLTGETGRLGQC-----GADACFPTNR-----TYLHHMTKAFIFYG 133  
DB 60 SIFLLIMTWPFKKFPLALPFCSENVPTADCNITADRKVYPQADAVYVHRE---VMYN 115  
QY 134 TDENIDSLP-LPRKAHHDAVVFHESPKNNYKLFHKFVITLFNVYATFSRSHLPLTTQY 192  
DB 116 PS---AQLPSPRROGQRWTFMESPSHCWQ--KAMDGFN-----LTMSY 158  
QY 193 LBSIEIVAKSLRYLVP-----LQSNKRLKRLAPLVYQSDCDPSPDRDSYREL 241  
DB 159 RQSDSDITFPYGMLEPWSGQPAHPPLNLSATEL----VAMVSNMGPNASARRYQSL 212  
QY 242 MTYIEVDSYGECLRNKDLPOOLKNPASMADGFRITIAQYKFLAFENAVACDDYITTEKFW 301  
DB 213 QALIKVDVYGR--SHKPLPQ-----GTMMETLSRKXFLAFENSLHPDITTKLW 260  
QY 302 R-ELKGVVFPVYVGSISITW---LPSNKAALVSESHRELASTYRRLDSDDLLEYAY 357  
DB 261 RNALLEMAVPAFVVLG-PSRSYVERFLPPD-AFIVHDVFOQKDLARYLQELDKDHARLASY 318  
QY 358 VEMKTLGEINQQLPALRER--KMGQDVNQNYIDAFECWCCTKWANIRLOEKL 413  
DB 319 FRRR-----ETLRPRSTSMAL-----AF-CKACWKLQESRFRYQTRGI 354

RESULT 4  
T23491  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T23491  
C:Smyle, R.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 21:39:06; Search time 42 Seconds  
(without alignments)  
1113.073 Million cell updates/sec

Title: US-10-080-960-2

Perfect score: 2625

Sequence: 1 MKVTGPPQGVTDMSMQCFNDQ.....LVDRNQNTSSQEPFGLVFKD 486

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283366 segs, 36191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 282.5 | 10.8        | 365    | 2 S55498  | alpha(1,3/4)-fucos |
| 2          | 279   | 10.6        | 414    | 2 T15270  | hypothetical prote |
| 3          | 263   | 10.0        | 359    | 2 A45156  | alpha-(1,3)-fucosy |
| 4          | 255.5 | 9.7         | 451    | 2 T23491  | hypothetical prote |
| 5          | 251   | 9.6         | 364    | 2 T39048  | alpha (1,3) fucosy |
| 6          | 245.5 | 9.4         | 374    | 2 A42270  | alpha (1,3) fucosy |
| 7          | 239.5 | 9.1         | 361    | 2 A36669  | 3-alpha-galactosyl |
| 8          | 230   | 8.8         | 393    | 2 H96742  | hypothetical prote |
| 9          | 224   | 8.5         | 342    | 2 A54057  | alpha(1,3)-fucosyl |
| 10         | 221   | 8.4         | 405    | 2 B36340  | alpha(1,3)-fucosyl |
| 11         | 213   | 8.1         | 433    | 2 A57596  | alpha-1,3-fucosyl  |
| 12         | 208   | 7.9         | 400    | 2 UC4591  | alpha-1,3-fucosyl  |
| 13         | 203   | 7.7         | 513    | 2 F96533  | probable fucosyltr |
| 14         | 185.5 | 7.1         | 304    | 2 T139049 | alpha (1,3) fucosy |
| 15         | 173.5 | 6.6         | 1652   | 2 T16799  | hypothetical prote |
| 16         | 144.5 | 5.5         | 346    | 2 T44327  | hypothetical prote |
| 17         | 143   | 5.4         | 425    | 2 C64567  | fucosyltransferase |
| 18         | 142   | 5.4         | 476    | 2 C64601  | fucosyltransferase |
| 19         | 136.5 | 5.2         | 436    | 2 G71862  | alpha-(1,3)-fucosy |
| 20         | 116   | 5.2         | 454    | 2 B71914  | alpha (1,3)-fucosy |
| 21         | 116   | 4.4         | 888    | 2 S50801  | AMP deaminase homo |
| 22         | 113   | 4.3         | 727    | 2 D86171  | hypothetical prote |
| 23         | 112.5 | 4.3         | 759    | 2 F85077  | probable transposo |
| 24         | 110   | 4.2         | 488    | 2 S64140  | hypothetical prote |
| 25         | 110   | 4.2         | 823    | 2 T09882  | heat shock protein |
| 26         | 106.5 | 4.1         | 3061   | 1 UN0545  | genome polypotein  |
| 27         | 106.5 | 4.1         | 3063   | 2 US0166  | genome polypotein  |
| 28         | 106   | 4.0         | 655    | 2 S57119  | hypothetical prote |
| 29         | 105.5 | 4.0         | 183    | 2 C97832  | alpha-(1,3)-fucosy |

|    |       |     |      |          |                    |
|----|-------|-----|------|----------|--------------------|
| 30 | 105.5 | 4.0 | 964  | 2 T01860 | reverse transcript |
| 31 | 104   | 4.0 | 817  | 2 S39558 | HSP90 homolog - Ma |
| 32 | 104   | 3.9 | 993  | 2 C31915 | lambdoidic nisin   |
| 33 | 103   | 3.9 | 979  | 2 T41006 | ubiquitin carboxyl |
| 34 | 103   | 3.9 | 2103 | 1 J01621 | genome polypotein  |
| 35 | 102.5 | 3.9 | 650  | 2 G70169 | heat shock protein |
| 36 | 101.5 | 3.9 | 2330 | 2 T25836 | hypothetical prote |
| 37 | 101   | 3.8 | 324  | 2 A41879 | hypothetical prote |
| 38 | 100.5 | 3.8 | 533  | 2 B26446 | vire2 protein - Ag |
| 39 | 100.5 | 3.8 | 789  | 2 G90587 | lipoprotein (lipor |
| 40 | 100.5 | 3.8 | 862  | 2 T38996 | hypothetical prote |
| 41 | 100.5 | 3.8 | 1411 | 2 S55123 | hypothetical prote |
| 42 | 100   | 3.8 | 842  | 2 D90240 | valyl-tRNA synthet |
| 43 | 100   | 3.8 | 2253 | 2 T30336 | nuclear/mitotic ap |
| 44 | 99.5  | 3.8 | 4196 | 2 T43274 | dynein heavy chain |
| 45 | 99    | 3.8 | 653  | 2 B81277 | hypothetical prote |

ALIGNMENTS

RESULT 1

S55498  
alpha(1,3/4)-fucosyltransferase - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 01-Aug-1995 #sequence\_revision 01-Sep-1995 #text\_change 13-Sep-1998  
C:Accession: S55498  
R:Olmuden, A.; Wierinckx, A.; Petit, J.M.; Julien, R.  
Submitted to the EMBL Data Library, June 1995  
A:Description: Molecular Cloning and expression of bovine alpha (1,3/4)-fucosyltransf  
A:Reference number: S55498  
A:Accession: S55498  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-365 <OUL>  
A:Cross-references: EMBL:X87810, NID:9860807, PID:9860808  
C:Superfamily: galactoside 3(4)-L-fucosyltransferase

Query Match 10.8%; Score 282.5; DB 2; Length 365;

Best local Similarity 28.4%; Pred. No. 3e-14; Matches 97; Conservative 49; Mismatches 117; Indels 79; Gaps 16;

|    |     |   |  |
|----|-----|---|--|
| QY | 90  | IMLWSPLTGETGRIGQC-----GADACFTTINFT-----YLH-----HMTKALFY 132   |  |
| DB | 69  | VILMTWPF-NQPVALSRCSELMPTGADQQLTVNSSEYQADAVFVHREVSHPKQL-- 125  |  |
| QY | 133 | GTDFTNIDSLPRKKAHDMVAFHESPPKNNKLFHKPVLTFENYATSPSHLPLTQY 192    |  |
| DB | 126 | -----PSPRPADQQRWVFMSPSPNCIKL--KIDGYFNLTMGTYRDSDFWYGM 175      |  |
| QY | 193 | LESIEVLKSLRYVLPGSKNLKRLKPLAVVYQSDCPSPDRSDSYVELMTYLEVDSYGE 252 |  |
| DB | 176 | LSPWP-SQVETRLINSAKTKL-----VAWVNSNMWTDLSRVQYKLLKPHLQVDVYGR 228 |  |
| QY | 253 | CIRKNDLPQQLKPNASMDGFRITIAQYKILAFENAVCDYITTECFMR-PLKLVPPV 311  |  |
| DB | 229 | F-HPTLPHAL-----MAKQLSQYKYLAFENSLHFDYITTEKLSNAQAQMAVPPV 276    |  |
| QY | 312 | YVGSPPIT--DWLPSNKALIVSESHPRELASYIRLSDSDRLYEAVVEWKLGEISNQ 369  |  |
| DB | 277 | VIGPRVNVQGLPP-KAFIHVEDQSPKDLAQYLLALDKQYAYLVNVPFMR----- 327    |  |
| QY | 370 | RLLTLALREKRGVQVDVNDONTIDAFECVNCYCTKVAANIRLQEK 411             |  |
| DB | 328 | ---ETLRPRSF-----SWALMFRCACW---KLQGE 351                       |  |

RESULT 2

T15270  
hypothetical protein F59312.13 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Mar-2000  
C:Accession: T15270

**This Page Blank (uspto)**

transferring a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine (generating the Lewis x or y epitope). It does not transfer a fucose moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine and therefore does not generate the sialyl-Lewis x epitope. The invention also relates to DNA sequences encoding alpha-1,3-fucosyltransferase and expression vectors and host cells comprising these DNA sequences. The invention additionally encompasses the preparation of alpha-1,3-fucosyltransferase via the culture of transformed cells or by expression of the protein in a transgenic animal; antibodies which recognise alpha-1,3-fucosyltransferase; methods for screening potential inhibitors or potentiators of alpha-1,3-fucosyltransferase activity or expression; the preparation of compounds having fucose-containing sugar chains by use of the protein; and knockout non-human animals lacking alpha-1,3-fucosyltransferase. Alpha-1,3-fucosyltransferase has a similar substrate range to the known FUC-III and is expressed mainly in brain and kidney tissues. Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies, potentiators and inhibitors may be used for the treatment and diagnosis of diseases of the brain and kidney, and of cancers. They may be used for the identification of substances which affect the activity or expression of alpha-1,3-fucosyltransferase; such substances may be used therapeutically. The knockout animals can be used to study the mechanisms of action and expression of alpha-1,3-fucosyltransferase. The present sequence represents human alpha-1,3-fucosyltransferase

Sequence 359 AA:

Query Match 10.5%; Score 275.5; DB 3; Length 359;  
Best Local Similarity 26.9%; Pred. No. 3.9e-19;  
Matches 99; Conservative 63; Mismatches 120; Indels 83; Gaps 18

QY 59 MEAPFHLNLFKKEGLFNRKRKELDSYIMLMSPLTGFGRIGQCA---DACEF 114  
Db 41 MESA---SSVAKMKNF-FSTKTDY-FNETTILWVWP-GQTFDDLTSQAMENIOGCHL 93  
QY 115 TIRRTTLHHMKAFLEFGTDPENIDSLPFRKAF---HDAVAFHEESPKNVYLFKFP-- 169  
Db 94 TIDRSL--YNSKHAVALIHRDISWDLTLPQQAAPPQKMIWNLSPT-----HTPQK 145  
QY 170 --VITLFNYATFSSHSHPLTTOYLESIEVLKSLRYLVLPSQKNKLKRLAFLVYQSD 227  
Db 146 SGIEHLFNLTLTYRSDSIQVPYGRL---TVSTNPVFVPSKEXL-----VCWVSN 195  
QY 228 CDPSPSRDSYVNEMLTYEVDSY---GECIRANDLPQQLKNSPMSDADSFYIILNQYF 283  
Db 196 WNEEHARVRYNELSKSLIHTYGOAFGEYVNDKSL-----IPTISACKF 240  
QY 284 ILAFENAVDDDYITEKFRPPLKLGVPVPYYSPS---ITDMLPSNKGALIVSEFSHPRL 340  
Db 241 YLSFENSIMKDIITTEKLIYNALAGSVPLVLS-PERENYENYITAD-SFIVEDYNSPSL 298  
QY 341 ASYIRRLSDDFLYEAYVWEKLGKGISNQRLTLIRKRGVQD--VNQDNYIDAECVY 398  
Db 299 AKYLKEVDKNNKLYLSYFNMR-----KDFVNLPRFWESHACIA 337  
QY 399 CTKY 402  
Db 338 CDHV 341

Search completed: September 9, 2004, 21:44:02  
Job time: 130 secs







OS Homo sapiens.  
XX Key Location/Qualifiers  
FH Peptide 1  
FT /label=Signal\_peptide  
FT Protein 2,261  
FT /label=mature\_secreted\_peptide  
XX MO200214341-A1.  
XX 21-FEB-2002.  
PD 13-AUG-2001; 2001MO-US025288.  
PF 14-AUG-2000; 2000US-0225215P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Barash SC, Ni J, Ruben SM, Rosen CA, Shi Y;  
XX MPI, 2002-257592/30.  
DR N-PSDB; ABK11711.  
XX Novel human secreted proteins useful for treating immune system e.g.  
PT anemia, cardiovascular e.g. myocardial infarction, neurological disorders  
PT e.g. Alzheimer's disease, Parkinson's disease.  
PS Claim 11; Page 330; 337pp; English.  
XX The invention describes eight novel human secreted polypeptides (I). (1)  
CC and the polynucleotides encoding them are useful for treating disorders,  
CC including disorders of immune system e.g. immunologic deficiency  
CC syndrome, ataxia telangiectasia, HIV infection (human immunodeficiency  
CC virus), Wiskott-Aldrich disorder, anaemia, thrombocytopenia, heart  
CC attacks (infarction), strokes, or scarring; autoimmune diseases e.g.  
CC Addison's disease, encephalomyelitis, multiple sclerosis, autoimmune  
CC thyroiditis, lupus, erythematosis, insulin dependent diabetes mellitus;  
CC disorders of inflammation e.g. nephritis, ischaemia-reperfusion injury,  
CC hypergammaglobulinaemia, Crohn's disease, sarcoidosis, Gaucher's disease;  
CC cardiovascular disease e.g. cardiovascular abnormalities, persistent  
CC truncus arteriosus, arrhythmias, angina pectoris, myocardial infarction,  
CC hypertension, telangiectasia, ischaemia includes cerebral ischaemia,  
CC mucocutaneous lymph node syndrome, disorders and/or conditions associated  
CC with neovascularisation; cancers which involve angiogenesis, benign  
CC tumours, ocular angiogenic diseases, hyperproliferative disorders,  
CC diabetic retinopathy, uveitis, rheumatoid arthritis, psoriasis, delayed  
CC wound healing, endometriosis, vasculogenesis, atherosclerosis; nervous  
CC system diseases, degeneration associated with Parkinson's disease,  
CC Alzheimer's disease and infectious diseases. This is the amino acid  
CC sequence of a novel human secreted protein, described in the invention  
XX Sequence 261 AA;  
Query Match 18.6%; Score 487.5; DB 5; Length 261;  
Best Local Similarity 39.4%; Pred. No. 3e-41;  
Matches 98; Conservative 51; Mismatches 99; Indels 11; Gaps 5;  
QY 242 MTYIEVDSYGECLEKNDLP-QQLKNPASMADG--FRITAGYKFTLAFENAVDDYITE 298  
DB 1 KRHIIVDSYGKCLQRELPRLADITATTEDEBELAFSRKXPHLALENAINCYDYTE 60  
QY 299 KFMRLKGVVYVYAGSPSTDM.LPSKSAITVEFHPHELASVTRRSDRLTEAVV 358  
DB 61 KLMRPMLHGAAPVVRGSPVRDMPNNHNVLLIDFSPKLAEFIDELKNDSEYKYL 120  
QY 359 EWLKGEISNORLLTALREKRGVQDVQNDYIDAFECMTCTKWMNIRLQE----KGL 413  
DB 121 AYKGGGIGTNGFILDLSLHREMGVNDPLPYLNGFCFCYDEELARLDKXHAASPGD 180  
QY 414 PP--KRWAEEDTHLSCEPPTVFAFSPPLTPLSLREMWISFEGSKKKAQALRMVVDN 471  
DB 181 SPVFEPHIAQSDHMDCPVTPFGFNVBERIPENDSWKEMWLODYQGLDGEALTAMIHNN 240

QY 472 QNFSQSEFW 480  
DB 241 ET-EQTKFW 248  
RESULT 12  
ID AU23474 standard; protein; 225 AA.  
XX AU23474;  
XX AAU23474;  
XX 18-DEC-2001 (first entry)  
XX DE Novel human enzyme polypeptide #560.  
XX Human; oxidoreductase enzyme; transferase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;  
KW nephrotropic; anticoagulant.  
XX  
OS Homo sapiens.  
XX MO200155301-A2.  
PN 02-AUG-2001.  
XX 17-JAN-2001; 2001MO-US001239.  
PF 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-019123P.  
PR 19-MAY-2000; 2000US-020515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.

XX 10-OCT-2002.  
 PD  
 XX  
 PF 17-JAN-2001; 2001US-00764877.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0106082P.  
 PR 28-JUN-2000; 2000US-0214856P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 11-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220363P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225417P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 21-SEP-2000; 2000US-0234233P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 XX  
 XX (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC,  
 XX  
 XX WPI; 2003-128199/12.  
 DR N-PSDB; ABX57833.  
 XX  
 PT Isolated nucleic acid molecules encoding musculoskeletal system  
 associated polypeptides, useful for detecting disorders, e.g. cancer.  
 PS Claim 11, SEQ ID NO 1210, 31pp. English.  
 XX  
 CC The invention describes an isolated nucleic acid molecule comprising a  
 CC sequence encoding musculoskeletal system associated polypeptides useful  
 CC for detecting disorders, e.g., cancer or cancer metastases, in animals or

CC humans. The nucleic acid: stimulates re-vascularisation of ischaemic  
 CC tissues associated with conditions such as thrombosis, arteriosclerosis,  
 CC and other cardiovascular conditions; treats wounds due to injuries;  
 CC burns, post-operative tissue repair, and ulcers; stimulates angiogenesis  
 CC and limb regeneration; stimulates neuronal growth; can treat and prevent  
 CC neuronal damage occurring in certain disorders or neurodegenerative  
 CC conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-  
 CC related complex; stimulates chondrocyte growth, thus they can be used to  
 CC enhance bone and periodontal regeneration and aid in tissue transports or  
 CC bone grafts; prevents skin aging due to sunburn by stimulating  
 CC keratinocyte growth; prevents hair loss, since RGF family members  
 CC activate hair forming cells and promotes melanocyte growth; stimulates  
 CC growth and differentiation of hematopoietic cells and bone marrow cells  
 CC when used in combination with other cytokines; maintains organs before  
 CC transplantation or for supporting cell culture of primary tissues;  
 CC induces tissue of mesodermal origin to differentiate in early embryos;  
 CC increases or decreases the differentiation or proliferation of embryonic  
 CC stem cells, besides, haematopoietic lineage; modulates mammalian  
 CC characteristics, such as, body height, weight, hair colour, eye colour,  
 CC skin, percentage of adipose tissue, pigmentation, size, and shape (e.g.,  
 CC cosmetic surgery); modulates mammalian metabolism; changes mammal's metal  
 CC state or physical state by influencing biorhythms, circadian rhythms,  
 CC depression, tendency for violence, tolerance for pain, reproductive  
 CC capabilities, hormonal or endocrine levels, appetite, libido, memory, or  
 CC stress; increases or decreases storage capabilities, fat content, lipid,  
 CC protein, carbohydrate, vitamins, minerals, cofactors or other nutritional  
 CC components. This is the amino acid sequence of a novel human  
 CC musculoskeletal system antigen. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from the US patent office at  
 CC ftp.segdata.uspto.gov/sequence.html?docID=200201471140  
 XX  
 SQ Sequence 107 AA:  
 Query Match 19.8%; Score 521; DB 6; Length 107;  
 Best Local Similarity 99.0%; Pred. No. 2.5e-45;  
 Matches 100; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 178 ATPRSRSHPLPTTYLLESLIEVLKSRIVVPQSKKRLKRLAPLVYQSDCDPPSDSDSY 237  
 DB 1 ATPRSRSHPLPTTYLLESLIEVLKSRIVVPQSKKRLKRLAPLVYQSDCDPPSDSDSY 60  
 QY 238 VEELMTYIEVDSYGECLELNKDLPOOLKKNPAMDADGFRIM 278  
 DB 61 VEELMTYIEVDSYGECLELNKDLPOOLKKNPAMDADGFRIM 101  
 RESULT 11  
 AAU7731  
 ID AAU7731 standard; protein; 261 AA.  
 XX  
 AC AAU7731;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE Novel human secreted protein #3.  
 XX  
 XX Secreted protein; gene therapy; immunostimulant; human; stroke;  
 KW immune system disorder; immunologic deficiency syndrome; anaemia;  
 KW ataxia telangiectasia; HIV; human immunodeficiency virus; heart attack;  
 KW Wiskott-Aldrich disorder; thymocyclopentia; scarring; erythematous;  
 KW autoimmune disease; Addison's disease; encephalomyelitis; lupus;  
 KW multiple sclerosis; autoimmune thyroiditis; diabetes; inflammation;  
 KW nephritis; ischaemia; hypergammaglobulinaemia; Crohn's disease;  
 KW sarcoidosis; Gaucher's disease; cardiovascular disease; telangiectasia;  
 KW persistent truncus arteriosus; arrhythmias; angina pectoris; tumour;  
 KW myocardial infarction; hypertension; neovascularisation; angiogenesis;  
 KW cancer; ocular angiogenic disease; hyperproliferative disorder;  
 KW diabetic retinopathy; uveitis; rheumatoid arthritis; psoriasis;  
 KW wound healing; endometriosis; vasculogenesis; atherosclerosis;  
 KW nervous system disease; Parkinson's disease; Alzheimer's disease;  
 XX infectious disease.  
 XX

|                           |  |
|---------------------------|--|
| XX                        | MP1: 2001-451937/48.   |
| DR                        | N-PsDB; AAJ34845.  |
| XX                        | Isolated polypeptide for treating, preventing and/or prognosing            |
| PT                        | disorders related to the musculoskeletal system including musculoskeletal  |
| PT                        | cancers and also for testing and detection e.g. diagnosis.                 |
| PS                        |  |
| XX                        | ClaIn 11; SEQ ID NO 1210; 781bp + Sequence Listing; English.               |
| CC                        | The invention relates to novel genes (AAJ34669-AAJ37666) and proteins      |
| CC                        | (ABB03087-ABB04109) associated with the musculoskeletal system useful for  |
| CC                        | preventing, treating or ameliorating medical conditions e.g. by protein    |
| CC                        | or gene therapy. The genes are isolated from a range of human tissues      |
| CC                        | disclosed in the specification. The nucleic acids, proteins, antibodies    |
| CC                        | and (ant)agonists are useful in the diagnosis, treatment and prevention    |
| CC                        | of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the    |
| CC                        | adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,   |
| CC                        | lung, or urogenital; (b) immune disorders e.g. Addison's disease,          |
| CC                        | allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,          |
| CC                        | diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid         |
| CC                        | arthritis and ulcerative colitis; (c) cardiovascular disorders such as     |
| CC                        | myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.   |
| CC                        | cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,   |
| CC                        | bacterial, fungal and parasitic infections. Note: The sequence data for    |
| CC                        | this patent did not form part of the printed specification, but was        |
| CC                        | obtained in electronic format directly from WIPO at                        |
| XX                        | ftp.wipo.int/pub/published_pct_sequences                                   |
| SQ                        | Sequence 107 AA;   |
| Query Match               | 19.8%; Score 521; DB 4; Length 107;  |
| Best Local Similarity     | 99.0%; Pred. No. 2,5e-45;  |
| Matches 100; Conservative | 1; Mismatches 0; Indels 0; Gaps 0  |
| YY                        | 178 ATFSRSHSLPITYTYSLEIEVYKSLRYVPVPIQSXKNKLREKLAPLVYQSDCPDPSDDSY 237       |
| Db                        | 1 AIFSRSHSLPLTYTYSLEIEVYKSLRIYVPIQSXKNKLREKLAPLVYQSDPDPSDDSY 60            |
| YY                        | 238 VAEIMTYIEVDSYGECILRNKDLPQOLKNPASMADGGFRII 278                          |
| Db                        | 61 VAEIMTYIEVDSYGECILRNKDLPQOLKNPASMADGGFRII 101                           |
| RESULT 10                 |  |
| ID                        | ABU12557 standard; protein; 107 AA.  |
| AC                        | ABU12557;  |
| XX                        | 26-FEB-2003 (first entry)  |
| DE                        | Novel human musculoskeletal system antigen #177.                           |
| XX                        | Musculoskeletal system antigen; cancer; metastasis; re-vascularisation;    |
| KM                        | thrombosis; arteriosclerosis; mineral content; cardiovascular condition;   |
| KM                        | wound; injury; burn; angiogenesis; ulcer; post-operative tissue repair;    |
| KM                        | limb regeneration; neuronal growth; neurodegenerative disorder;            |
| KM                        | Alzheimer's disease; Parkinson's disease; AIDS-related complex;            |
| KM                        | chondrocyte growth; bone regeneration; periodontal regeneration;           |
| KM                        | tissue transport; bone graft; skin aging; keratinocyte growth; hair loss;  |
| KM                        | melanocyte growth; cell proliferation; cell growth; organ transplant; skin |
| KM                        | cell differentiation; body height; weight; hair colour; eye colour; skin   |
| KM                        | percentage of adipose tissue; pigmentation; cosmetic surgery; metabolism;  |
| KM                        | biorythm; cardiac rhythm; depression; tendency for violence; pain;         |
| KM                        | reproductive capability; hormone level; endocrine level; appetite;         |
| KM                        | libido; memory; stress; storage capability; fat content; lipid content;    |
| KM                        | protein content; carbohydrate content; vitamin content; cofactor content   |
| XX                        | nutritional component.   |
| CS                        | Homo sapiens.  |
| XX                        | US2002147140-A1.   |

[illegible]

|    |  |              |             |            |             |
|----|--|--------------|-------------|------------|-------------|
|    | Query Match  | 27.1%        | Score 710.5 | DB 4       | Length 443  |
|    | Best Local Similarity                                      | 37.2%        | Pred. No.   | 6.3e-64    |             |
|    | Matches 156;   | Conservative | 81;         | Mismatches | 153;        |
|    |  |              |             | Indels     | 29; Gaps 10 |
| OY | 68 SLKSGGLFNRKKRWELDSPI-MLWMS-PVLTGTCRLGGCGCADFFINTITYHHHM | 125          | :           | : : :      | :           |

```

RESULT 8
ABG92064
ID   ABG92064 standard; protein; 168 AA.

```

DT 29-NOV-2002 (first entry)

KM Human receptor and membrane-associated protein REMP; atherosclerosis,  
 cardiovascular disorder; hypertension, congestive heart failure; oedema,  
 KM aneurysm; angina pectoris; ischaemic heart disease; lung disease; nausea,  
 KM rheumatic heart disease; chronic obstructive pulmonary disease; anaemia;  
 KM emphysema; bronchitis; gastrointestinal disorder; peptic ulcer; actinia;  
 KM Crohn's disease; lipid metabolism disorder; Fabry's disease; dementia;  
 KM pancreatitis; diabetes mellitus; hyperlipidaemia; autoimmune disorder;  
 KM inflammatory disorder; neurological disorder; Creutzfeldt-Jakob disease;  
 KM cell proliferative disorder; protein replacement therapy; adenocarcinoma;  
 KM developmental disorder; metabolic disorder; Alzheimer's disease; stroke;  
 KM Parkinson's disease; multiple sclerosis; Cushing's syndrome; anxiety;  
 KM schizophrenia; Addison's disease; endocrine disorder; gene therapy; gout;  
 KM amnesia; cancer; leukaemia; lymphoma; myeloma; sarcoma.

PN WO200263006-A2

PF 05-FEB-2002; 2002WO-US003868.

PR 06-FEB-2001; 2001US-0267201P.

PR 09-APR-2001; 2001US-0282679P.

PR 14-JAN-2002; 2002US-0348687P.

PA (INCY-) INCYTE GENOMICS INC.  
XX

P1 Azamzi Y, Yue H, Ding L, Nguyen DB, Gandhi AR, Burford N;  
P1 Thangavelu K, Elliott VS, Ramkumar Y, Yao MG, Lal PG, Tang TY;  
P1 Swarnikar A, Warren BA, Walla NK, Pollock JL, Xu Y, Honchell CD;  
P1 Au-Yang J, Baugh MR, Duggan BM, Lu DM, Gleason KJ, Hillman JL;  
P1 Rammann BE, Lu Y, Saprestein SK, Tran UC, Richardson TW;  
P1 Emerling BM, Hafila AOA, Buttrill JD, Marcus GA, Zingler KA;

CC member of the fucosyltransferase family comprising: (a) nucleotide  
 CC sequence having 80% identity to sequence S1 (AB05333) or S3 (AB05334);  
 CC (b) comprising fragment of S1 or S3; (c) encoding a polypeptide (II)  
 CC comprising a fully defined sequence S2 (AA047905) of 492 amino acids or  
 CC its fragment; or (d) encoding naturally occurring allelic variant of S2.  
 CC (1) has cytostatic, antitumor, antidiabetic, antihypertensive,  
 CC antiarhythmic, vasotropic, antidiabetic, antitachycardic,  
 CC immunosuppressive, antihypertensive, antitumor, antidiabetic,  
 CC ophthalmological, antiallergic, antiaesthetic, nephroprotective,  
 CC dermatological, antidiarrhetic, vulnerary, haemostatic, virucide,  
 CC antibacterial, neurotropic, neuroprotective, antiparkinsonian,  
 CC anticonvulsant, anesthetic, anorectic, metabolic and immunomodulator  
 CC activity. (I) and (II) are useful for treating proliferative and/or  
 CC differentiative disorders of the colon, including adenoma and colorectal  
 CC carcinogenesis, the liver, including nodular hyperplasia and adenomas,  
 CC breast, including epithelial hyperplasia and sclerosing adenosis, lung,  
 CC including bronchogenic carcinoma and neuroendocrine tumor. The  
 CC proliferative disorders treated also include haematopoietic neoplastic  
 CC disorders such as acute promyeloid leukaemia, acute myelogenous  
 CC leukaemia. (I) and (II) are also useful for treating cardiovascular  
 CC diseases, disorders of blood vessels, immunological disorders such as  
 CC diabetes mellitus, arthritis, ulcerative colitis, Crohn's disease,  
 CC asthma, kidney disorders e.g., polycystic kidney disease, acute  
 CC glomerulonephritis, colon disorders, disorders of liver such as disorders  
 CC associated with accumulation of fibrous tissue, hepatocellular necrosis  
 CC or injury induced by agents including processes which disturb  
 CC homeostasis, disorders of brain such as intracranial haemorrhage,  
 CC bacterial or viral meningitis, neurodegenerative disorder, glioma,  
 CC multiple sclerosis, metabolic or pain disorders. (I) can be used in gene  
 CC therapy  
 CC XX

Sequence 492 AA;

Query Match 33.2%; Score 872; DB 5; Length 492;

Best Local Similarity 42.6%; Pred. No. 1.5e-80;

Matches 178; Conservative 74; Mismatches 146; Indels 20; Gaps 9;

QY 79 RRRKKWLDSEYIMLWMSF-----LJGEGRLGCGCADAQFTIRTYIHMMETAFIFYG 133  
 DB 66 RGRERAGDLPVLLWMSFGLPFPHPEDSERI-ECARGACVASRRRPARDSRTALPYG 124  
 QY 134 TDENIDSLPFRKAHDMVAFHEESPKNVYKLFHKEVITLTFNYATFSRSHLPLTQYL 193  
 DB 125 TDFRASAAPLPRLAHQSMALHESBPNNFLSHGPGIRLFNLTSTFSRSHDYLTLQWL 184  
 QY 194 ESTEVLKSLRYLVLDSKKKLR-LAPLVYVQSCDPPSPDSVYELMTYIVDSYGS 252  
 DB 185 PGTAAYLR-RPVPPPERAEWRRRGAPLTLQSHCDVPADRDRVRLMHIIVDSYGS 242  
 QY 253 CLRNKSLP-QOLKNPMSMDAG--FYRIIAQYKFLAFENAVCDYIEKFWPLTKGVV 309  
 DB 243 CLONRLPFRARLQDTATTEDEPELLALSKRYKFLAENALCNDYMKRMKRMHGAIV 302  
 QY 310 PYYGSPSITDMLPSNKAIVLSEFSHPRELAAYIRLSDRLYEAVEMKLGKISNQ 369  
 DB 303 PYYRGSPVYRDMWNNHVSILLIDDESFOKLAEFIDFKKDEEMKYLAVKOGGITNQ 362  
 QY 370 RLTLALRRKMGVQDQVQNDNYIDAFECWCTKVMANILQF-----KSLRP--KRWAEAD 422  
 DB 363 FLDDSLKREKGVNDPLPNYNGECFCVDEYELRLAEKAAASPDSPVFPFPIAQP 422  
 QY 423 TLSCDEPTVFAFSLRTPPLSLREMMISSEFGSKQAQALRWLVDRNQNFSSQEFW 480  
 DB 423 SHMDCPVPTPGFNGVVEIPENDSWKEMVLQDYWQGLDQEGALTLAMHNNET-EQTKFW 479

RESULT 6  
 ADCC32961  
 ID ADCC32961 standard; protein; 148 AA.  
 AC ADCC32961;  
 XX  
 XX 18-DEC-2003 (first entry)

XX Human novel contig-encoded polypeptide sequence, SEQ ID NO:3043.  
 DE Human, diagnostic; drug screening; forensics; gene mapping;  
 XX biodefensibility assessment; Parkinson's disease; Alzheimer's disease;  
 KM neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
 KM ulcers; osteoporosis; autoimmune disease; cancer;  
 KM molecular weight marker; food supplement; antiparkinsonian; neurotropic;  
 KM neuroprotective; antianemic; anticoagulant; thrombolytic; vulnerary;  
 KM anticancer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
 KM gene therapy; chromosome 8.  
 XX Homo sapiens.  
 OS  
 XX  
 XX MO2003029271-A2.  
 XX  
 XX 10-APR-2003.  
 XX  
 XX 24-SEP-2002; 2002MO-US030474.  
 XX  
 XX 24-SEP-2001; 2001US-0324631P.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
 PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
 PI Haley-Vicente D, Dmanac RT;  
 XX  
 XX WPI; 2003-371981/35.  
 DR N-PSDB; ADC32194.  
 XX  
 XX New polynucleotide and polypeptide useful for diagnosing, preventing or  
 PT treating conditions such as neurodegenerative diseases, anemias, platelet  
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 PT cancer.  
 XX  
 XX Example 2; SEQ ID NO 3043; 1185bp; English.

PS The invention relates to 971 novel human cDNA sequences (ADC2919P-  
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
 CC invention also relates to nucleic acid sequences over 99% identical with  
 CC the novel human cDNAs. The invention additionally encompasses expression  
 CC vectors and host cells comprising a nucleic acid of the invention; the  
 CC recombinant production of a polypeptide of the invention; an antibody  
 CC against a polypeptide of the invention; a method of detecting  
 CC polynucleotides or polypeptides of the invention; and methods of  
 CC identifying a compound which binds to a polypeptide of the invention. The  
 CC invention further discloses methods of preventing, treating or  
 CC ameliorating a medical condition, kits comprising polynucleotide probes  
 CC and/or monoclonal antibodies for carrying out the methods of the  
 CC invention; methods for the identification of compounds that modulate the  
 CC expression or activity of the polynucleotide and/or polypeptide; and 767  
 CC config. sequences corresponding to the cDNA sequences of the invention  
 CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628  
 CC -ADC33394). The nucleic acids and polypeptides of the invention are  
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
 CC identification of mutations responsible for genetic disorders or other  
 CC traits, for assessing biodiversity, and in producing many other types of  
 CC data and products dependent on DNA and amino acid sequences. They are  
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, anaemia, platelet  
 CC disorders, wounds, ulcers, osteoporosis, autoimmune diseases or  
 CC cancer. The nucleic acids may also be used as hybridisation probes or  
 CC primers, and in the recombinant production of a protein. The polypeptides  
 CC are also useful in generating antibodies, as molecular weight markers,  
 CC and as food supplements. The present sequence represents a human contig-  
 CC encoded polypeptide sequence used in an example of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 148 AA;

DT 08-JUL-2002 (first entry)  
XX  
DE Human ORF553 protein, SEQ ID NO:7106.  
XX  
KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
KW immune modulation; haematopoiesis regulation; tissue growth;  
KW angiogenesis; activin; inhibin; chemotactic; chemokine; haemostatic;  
KW thrombolytic; tumour inhibition; bodily characteristics; fertility;  
KW behaviour; cancer; proliferative disorder; neurological disorder;  
KW cardiovascular disease; immune system disorder; organ transplantation;  
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
KW hypothyroidism; cholesterol ester storage disease; infection; viraemia;  
KW vasculopathy; antidiabetic; antidiabetic; cytosolic; neotropic;  
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
KW carotid; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
KW dermatological; analgesic; virucide; antibacterial; fungicide.  
XX  
OS Homo sapiens.  
XX  
PN WO200190366-A2.  
XX  
XX  
XX 29-NOV-2001.  
XX  
XX 24-MAY-2001; 2001WO-US017076.  
XX  
XX 24-MAY-2000; 2000US-0206690P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Leach MD, Shinkets RA;  
XX  
XX WPI: 2002-106200/14.  
XX  
XX N-PSDB; AEN78606.  
XX  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
XX preventing and treating cardiovascular disease, neurodegenerative,  
XX hyperproliferative disorders and disorders related to organ  
XX transplantation.  
XX  
XX Claim 10; Page 2022; 2508bp; English.  
XX  
XX Sequences ABP31028-ABP3561 represent 4534 novel human proteins  
XX designated ORF (open reading frame) 1-4534, and sequences AEN75054-  
XX AEN79587 represent cDNAs encoding them. The invention also encompasses  
XX polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
XX referred to as ORFX) proteins, polynucleotides at least 85% identical to  
XX the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
XX polynucleotides, the recombinant production of ORFX proteins, antibodies  
XX specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
XX polypeptides, methods of screening for modulators of ORFX expression or  
XX activity, and methods of screening individuals for a predisposition to an  
XX ORFX-associated disorder. The ORFX proteins of the invention have a wide  
XX range of biological activities, such as cytokine, cell proliferation,  
XX cell differentiation, immune modulation, haematopoiesis regulation,  
XX tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
XX chemokinetic activity, haemostatic activity, thrombolytic activity,  
XX receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
XX and antiinfective activity, and may also be involved in the determination  
XX of bodily characteristics, fertility and behaviour. ORFX proteins,  
XX nucleic acids and antibodies may be used in the treatment of cancers,  
XX other proliferative disorders such as psoriasis and benign tumours,  
XX neurological disorders such as epilepsy and Alzheimer's disease,  
XX cardiovascular diseases, immune system disorders, disorders related to  
XX organ transplantation, disorders of tissue growth and regeneration,  
XX diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
XX storage disease, and infectious diseases caused by viral, bacterial,  
XX fungal and other pathogens. ORFX nucleic acids may also be used as a  
XX source of primers and probes, in the detection of ORFX genomic sequences  
XX or transcripts, in the identification and cloning of homologous  
XX sequences, in genetic diagnosis, and in forensic biology. The ORFX  
XX nucleic acids may additionally be used to produce transgenic animals  
XX which may be useful for studying the function and/or activity of ORFX

CC protein, and in drug screening. The ORFX proteins may also be used as  
CC immunogens to generate specific antibodies, which are useful in the  
CC diagnosis, treatment and monitoring of ORFX-associated diseases  
XX  
XX Sequence 197 AA;  
XX  
XX Query Match 38.7%; Score 1015; DB 5; Length 197;  
XX Best Local Similarity 99.0%; Pred. No. 6, 2e-96;  
XX Matches 197; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX 159 PKNVYKLFKHPVITLLENNYATFSRSHLPLTQVLESTIEVLSKLEYLPLOSXNLRRL 218  
XX |  
XX 3 PENNYKLFKHPVITLLENNYATFSRSHLPLTQVLESTIEVLSKLEYLPLOSXNLRRL 62  
XX |  
XX 219 APVYVQSDCDPSDRSDSVREIMTYIEVDSYGECLRNKDLPQQLKNPAMDADGFYRII 278  
XX |  
XX 63 APLVYQSDCDPSDRSDSVREIMTYIEVDSYGECLRNKDLPQQLKNPAMDADGFYRII 122  
XX |  
XX 279 AQRKFLAFENAVCDYITTEKFWRPKLGVPPVYVYGSF8ITWLESNKSAIIYVSEF8R 338  
XX |  
XX 123 AQRKFLAFENAVCDYITTEKFWRPKLGVPPVYVYGSF8ITWLESNKSAIIYVSEF8R 182  
XX |  
XX 339 ELASYIRRLDSDRL 353  
XX |  
XX 183 ELASYIRRLDSDRL 197  
XX  
XX  
XX RESULT 5  
XX AAM47905  
XX ID AAM47905 standard; protein; 492 AA.  
XX  
XX AAM47905;  
XX  
XX 26-FEB-2002 (first entry)  
XX  
XX  
XX Human fucosyltransferase family member 32132.  
XX  
XX Human; fucosyltransferase family member 32132; cytosolic; antitumour;  
XX antiatherosclerotic; hypotensive; antiarrhythmic; vasotropic;  
XX antidiabetic; antierectile; immunosuppressive; antithyroid; antiulcer;  
XX antiinflammatory; ophthalmological; antiallergic; antiasthmatic;  
XX nephrotoxic; dermatological; antidiarrhetic; viraemia; haemostatic;  
XX vitruclide; antibacterial; neotropic; neuroprotective; antiparkinsonian;  
XX anticonvulsant; analgesic; anorectic; metabolic; immunomodulator;  
XX proliferation; differentiation; cancer; apoptosis; leukaemia;  
XX cardiovascular disease; diabetes mellitus; ulcerative colitis;  
XX Crohn's disease; glomerulonephritis; hepatocellular necrosis;  
XX homeostasis; meningitis; multiple sclerosis; pain; gene therapy.  
XX  
XX Homo sapiens.  
XX  
XX WO200183721-A2.  
XX  
XX 08-NOV-2001.  
XX  
XX 27-APR-2001; 2001WO-US013805.  
XX  
XX 28-APR-2000; 2000US-0200604P.  
XX  
XX (MILL-) MILLBENTUM PHARM INC.  
XX  
XX Meyers RA, Williamson M;  
XX  
XX WPI: 2002-041492/05.  
XX  
XX N-PSDB; ABA05333, ABA05334.  
XX  
XX Novel isolated 32132 nucleic acid encoding 32132 polypeptide, a novel  
XX member of fucosyltransferase family, useful for treating atherosclerosis,  
XX multiple sclerosis, acute glomerulonephritis, Raynaud's disease, pain.  
XX  
XX Claim 9; Page 107; 125bp; English.  
XX  
XX The invention relates to an isolated 32132 nucleic acid (I), encoding a

CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
CC cancer. The nucleic acids may also be used as hybridization probes or  
CC primers, and in the recombinant production of a protein. The polypeptides  
CC are also useful in generating antibodies, as molecular weight markers,  
CC and as food supplements. The present sequence represents a specifically  
CC claimed human polypeptide sequence of the invention. Note: The sequence  
CC data for this patent did not form part of the invention. Note: The sequence  
CC was obtained in electronic format directly from WIP0 at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX

Sequence 546 AA:

Query Match 93.0%; Score 2442; DB 7; Length 546;  
Best Local Similarity 99.8%; Pred. No. 1,2e-242;  
Matches 453; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 33 KEVVEELGKPERKEFKSSSLQDGHTEMEAPTHLNSFLKKEGLTEMRKKEWELDSYPIML 92  
Db 93 KNNVVELGKPERKEFKSSSLQDGHTEMEAPTHLNSFLKKEGLTEMRKKEWELDSYPIML 152  
QY 93 WMSPLTGTGRLGCGGADACFTTINRTYLNHNTYAFLEFYGTDENIDSLPLPRKAHDWA 152  
Db 153 WMSPLTGTGRLGCGGADACFTTINRTYLNHNTYAFLEFYGTDENIDSLPLPRKAHDWA 212  
QY 153 VFHEESPKNKYKLFHKPVITLFFNYATFSRSHLPLTTQYLESIEVLKSLRYLVPLQSKN 212  
Db 213 VFHEESPKNKYKLFHKPVITLFFNYATFSRSHLPLTTQYLESIEVLKSLRYLVPLQSKN 272  
QY 213 KLKRLAFLVYVQSDCDPPSDRDSYRELMTYIEVDSYGECLRNKDLPOQLKNPASMAD 272  
Db 273 KLKRLAFLVYVQSDCDPPSDRDSYRELMTYIEVDSYGECLRNKDLPOQLKNPASMAD 332  
QY 273 GFRRIIAQYKFLAENAVCDYITEKFRPLKLGIVVYVQSPBITWMLSNKSATILVS 332  
Db 333 GFRRIIAQYKFLAENAVCDYITEKFRPLKLGIVVYVQSPBITWMLSNKSATILVS 392  
QY 333 EFSPRELASYIRLDSDDLRYEAYEWMLKGEISNOBLLTALRRKMGVQDVNQDYID 392  
Db 393 EFSPRELASYIRLDSDDLRYEAYEWMLKGEISNOBLLTALRRKMGVQDVNQDYID 452  
QY 393 AFECAWCTKWMANIRLOEGCLPPKWEADTHLSCPEPTVAFSPILRTPPLSSIREMWTIS 452  
Db 453 AFECAWCTKWMANIRLOEGCLPPKWEADTHLSCPEPTVAFSPILRTPPLSSIREMWTIS 512  
QY 453 SFEOSSKEAOLRWLVDRNQNPSSOEFMGLVFKD 486  
Db 513 SFEOSSKEAOLRWLVDRNQNPSSOEFMGLVFKD 546

#### RESULT 3

AAAG89164 standard; protein; 406 AA.

AC AAG89164;  
XX Homo sapiens.  
XX OS  
XX PN WO200142451-A2.  
XX PD 14-JUN-2001.  
XX PF 07-DEC-2000; 2000MO-IB001938.  
XX PR 08-DEC-1999; 99US-0169629P.  
XX PR 06-MAR-2000; 2000US-0187470P.

PA (GENT ) GENSER.

XX Dumas Milne Edwards J, Bougueleret L, Jobert S;

PI WPI: 2001-367870/38.

DR N-PSDB; AA64767.

XX Full length GENSER human nucleic acids encoding potentially secreted  
PT proteins, useful in gene therapy and vaccination against a variety of  
PT diseases, and for diagnosis of those diseases.  
XX

PS Claim 21; Page 819-820; 921pp; English.

CC The invention relates to full length GENSER human nucleic acids encoding  
CC potentially secreted proteins. The nucleic acids and the polypeptides  
CC they encode may be used in the prevention, treatment and diagnosis of  
CC diseases associated with inappropriate GENSER gene expression. For  
CC example, they be used to treat disorders associated with decreased GENSER  
CC gene expression by rectifying mutations or deletions in a patient's  
CC genome that affect the activity of GENSER or by supplementing the  
CC patient's own production of GENSER polypeptides. Conversely, antisense  
CC nucleic acid molecules may be administered to down regulate GENSER  
CC expression by binding with the cells' own genes and preventing their  
CC expression. The sense and antisense nucleic acids may also be used as  
CC probes in diagnostic assays to detect and quantitate the presence of  
CC similar nucleic acid sequences in samples, and hence to determine which  
CC patients may be in need of restorative therapy. The GENSER polypeptides  
CC may be used as antigens in the production of antibodies and in assays to  
CC identify modulators (agonists and antagonists) of GENSER polypeptide  
CC expression and activity. The present sequence is a GENSER polypeptide of  
CC the invention  
XX

Sequence 406 AA;

Query Match 74.6%; Score 1957.5; DB 4; Length 406;  
Best Local Similarity 97.9%; Pred. No. 8.4e-193;  
Matches 366; Conservative 1; Mismatches 2; Indels 5; Gaps 1;

QY 1 KCVTGPQGVDSMOCFNDQWPLNTRSSSEIKEWELGKFERKEFKSSSLQDGHTEME 60  
Db 36 KCVTGPQGVDSMOCFNDQWPLNTRSSSEIKEWELGKFERKEFKSSSLQDGHTEME 95  
QY 61 EAPTHLNSFLKKEGLTEFRKRWELDSYPIMLWMSPLTGTGRLGCGGADACFTTINRTY 120  
Db 96 EAPTHLNSFLKKEGLTEFRKRWELDSYPIMLWMSPLTGTGRLGCGGADACFTTINRTY 155  
QY 121 LHHNTKAFLEFGVDENIDSLPLPRKAHDWAVFHEESPKNKYKLFHKPVITLFFNYATF 180  
Db 156 LHHNTKAFLEFGVDENIDSLPLPRKAHDWAVFHEESPKNKYKLFHKPVITLFFNYATF 215  
QY 181 SRHSHLPLTTQYLESIEVLKSLRYLVPLQSKNKLKRLAFLVYVQSDCDPPSDRDSYVRE 240  
Db 216 SRHSHLPLTTQYLESIEVLKSLRYLVPLQSKNKLKRLAFLVYVQSDCDPPSDRDSYVRE 275  
QY 241 LMTYIEVDSYGECLRNKDLPOQLKNPASMADGFRRIIAQYKFLAENAVCDYITEKF 300  
Db 276 LMTYIEVDSYGECLRNKDLPOQLKNPASMADGFRRIIAQYKFLAENAVCDYITEKF 335  
QY 301 WRPLKLGIVVYVQSPITDWLPENKSAIIVSEFSPRELASYIRLDSDDLRYEAYEW 360  
Db 336 WRPLKLGIVVYVQSPITDWLPENKSAIIVSEFSPRELASYIRLDSDDLRYEAYEW 395  
QY 361 KLKGEISNOBLLTA 374  
Db 396 KLKGEISNOBLLTA 404

#### RESULT 4

ABP34580 standard; protein; 197 AA.

AC ABP34580;  
XX





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 21:30:50 / Search time 127 Seconds  
(without alignments)  
1081.245 Million cell updates/sec

Title: US-10-080-960-2  
Perfect score: 2625  
Sequence: 1 MKVTGPQGVTDMSQCFNDQ.....LVDRNQNFSSQFVGLVFXD 486

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

|    |                 |
|----|-----------------|
| 1: | geneseqp19806:* |
| 2: | geneseqp19908:* |
| 3: | geneseqp20008:* |
| 4: | geneseqp20018:* |
| 5: | geneseqp20028:* |
| 6: | geneseqp20038:* |
| 7: | geneseqp20038:* |
| 8: | geneseqp20045:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description        |
|------------|--------|-------------|--------|-------|--------------------|
| 1          | 2611   | 99.5        | 486    | 5     | AAU97908 Human 800 |
| 2          | 2442   | 93.0        | 546    | 7     | ADG31278 Human nov |
| 3          | 1957.5 | 74.6        | 406    | 4     | AA689164 Human sec |
| 4          | 1015   | 38.7        | 197    | 5     | ADP34580 Human ORF |
| 5          | 872    | 33.2        | 492    | 5     | AAm47905 Human ORF |
| 6          | 725    | 27.6        | 148    | 7     | ADG32961 Human nov |
| 7          | 710.5  | 27.1        | 443    | 4     | ABG60645 Human nov |
| 8          | 554.5  | 21.1        | 168    | 5     | ABG92064 Human nov |
| 9          | 521    | 19.8        | 107    | 4     | ABG03263 Human rec |
| 10         | 521    | 19.8        | 107    | 6     | ABU12557 Human mus |
| 11         | 487.5  | 18.6        | 261    | 5     | AAU77731 Novel hum |
| 12         | 485.5  | 18.5        | 225    | 4     | AAU23474 Novel hum |
| 13         | 279.5  | 10.6        | 525    | 4     | ABG62750 Novel hum |
| 14         | 278.5  | 10.6        | 359    | 3     | AAV80995 Drosophil |
| 15         | 275.5  | 10.5        | 359    | 3     | AAV80996 Murine al |
| 16         | 266    | 10.1        | 359    | 2     | AAW14524 Human alp |
| 17         | 265    | 10.1        | 359    | 2     | AAW14523 Human chi |
| 18         | 263    | 10.0        | 359    | 2     | AAW14522 Human chi |
| 19         | 262    | 10.0        | 359    | 2     | AAW14521 Human chi |
| 20         | 259.5  | 9.9         | 360    | 2     | AAW14520 Human chi |
| 21         | 258    | 9.8         | 360    | 2     | AAW14519 Human chi |
| 22         | 258    | 9.8         | 360    | 2     | AAW14518 Human chi |
| 23         | 258    | 9.8         | 360    | 2     | AAW14517 Human chi |
| 24         | 258    | 9.8         | 360    | 2     | AAW14516 Human chi |

|    |       |     |     |   |                    |
|----|-------|-----|-----|---|--------------------|
| 26 | 255   | 9.7 | 359 | 2 | AAW14525 Human chi |
| 27 | 254.5 | 9.7 | 360 | 2 | AAW14513 Human chi |
| 28 | 251.5 | 9.6 | 360 | 2 | AAW14512 Human chi |
| 29 | 251   | 9.6 | 360 | 2 | AAW14514 Human chi |
| 30 | 248   | 9.4 | 360 | 2 | AAW14515 Human chi |
| 31 | 247.5 | 9.4 | 361 | 2 | AAW14520 Human chi |
| 32 | 245.5 | 9.4 | 374 | 2 | AAW14539 Human alp |
| 33 | 245.5 | 9.4 | 374 | 2 | AAW13642 Human alp |
| 34 | 244.5 | 9.3 | 361 | 2 | AAW14530 Human chi |
| 35 | 244   | 9.3 | 350 | 2 | AAW14532 Human chi |
| 36 | 244   | 9.3 | 360 | 2 | AAW14516 Human chi |
| 37 | 243   | 9.3 | 360 | 2 | AAW14526 Human chi |
| 38 | 242.5 | 9.2 | 361 | 2 | AAW14527 Human chi |
| 39 | 240.5 | 9.2 | 361 | 2 | AAW14517 Human chi |
| 40 | 239.5 | 9.1 | 361 | 2 | AAW13749 Human chi |
| 41 | 239.5 | 9.1 | 361 | 2 | AAW45934 Human alp |
| 42 | 239.5 | 9.1 | 361 | 2 | AAW23806 Human alp |
| 43 | 239.5 | 9.1 | 361 | 2 | AAW13638 Human alp |
| 44 | 239.5 | 9.1 | 361 | 2 | AAW64452 Human alp |
| 45 | 237.5 | 9.0 | 361 | 2 | AAW14518 Human chi |

ALIGNMENTS

|          |   |   |
|----------|---|---|
| RESULT 1 | AAU97908  | standard; protein; 486 AA.                      |
| ID       | AAU97908  |   |
| XX       | AAU97908;   |   |
| AC       | AAU97908;   |   |
| XX       | 13-AUG-2002 (first entry)   |   |
| DT       | 13-AUG-2002   |   |
| XX       | Human 80090 protein.  |   |
| DE       | Human 80090 protein.  |   |
| XX       | Human; 80090; fucosyl transferase; cancer; Grave's disease;           |   |
| XX       | seven transmembrane receptor; Rho GTPase-activating protein; Rhodap;  |   |
| KW       | cardiovascular disorder; hematopoietic disorder; brain disorder;      |   |
| KW       | blood vessel disorder; metabolic disorder; liver disorder; psoriasis; |   |
| KW       | platelet disorder; leukemia; cardiac hypertrophy; Crohn's disease;    |   |
| KW       | myocardial infarction; rheumatic fever; multiple sclerosis; asthma;   |   |
| KW       | rheumatoid arthritis; diabetes mellitus; blood vessel disorder;       |   |
| KW       | Kaposi sarcoma; Alzheimer's disease; hormonal disorder.               |   |
| XX       |   |   |
| XX       | Homo sapiens.   |   |
| OS       | Homo sapiens.   |   |
| XX       |   |   |
| FH       | Key   | Location/Qualifiers                             |
| FT       | Domain  | 1..305  |
| FT       | Domain  | /note= "N terminal cytoplasmic domain"          |
| FT       | Domain  | 35..395   |
| FT       | Domain  | /note= "fucosyltransferase domain"              |
| FT       | Domain  | 82..265   |
| FT       | Domain  | /note= "CG435"                                  |
| FT       | Domain  | 90..390   |
| FT       | Domain  | /note= "CG169 domain"                           |
| FT       | Domain  | 221..367  |
| FT       | Domain  | /note= "Transferase fucosyltransferase          |
| FT       | Domain  | glycosyltransferase alpha-12-fucosyltransferase |
| FT       | Domain  | galactosidase domain"                           |
| PD       | WO200232962-A2.   |   |
| XX       | 25-APR-2002.  |   |
| XX       | 19-OCT-2001; 2001WO-US048307.   |   |
| XX       | 20-OCT-2000; 2000US-0241992P.   |   |
| XX       | 20-OCT-2000; 2000US-0242038P.   |   |
| XX       | 20-OCT-2000; 2000US-0242040P.   |   |
| XX       | 23-OCT-2000; 2000US-0242637P.   |   |
| XX       | (MILL-) MILLENNIUM PHARM INC.   |   |

```

Db      173  RSDSDIFTPYGMLEPWSGQPAHPPLNLAKTEL-----VAMAVSNWKPPDSARVRYQSL 226
QY      242  MTYIEVDSYGECLRNKDLPQOLKNVPASMDADGFYRIIAQYKFIAPENAVCDDYITEKFW 301
Db      227  QAHKAVDYGR--SHKPLPK-----CTMETLSRYKFIAPFENSLHPDYITEKLM 274
QY      302  R-PLKLGVPYVYSGPSITDW--LPSNKSAILVSEFSPRELASTYIRRLDSDDRLYEAY 357
Db      275  RNALBAMAVPYVLG-PSRSNYERFLPPD-AFIHVDDFQSPKDLARYLOELDKDHARYLSY 332
QY      358  VEWKLKGELISNQRLLTALRER--KMGVQDVNODNYIDAPECWVCTKVMANIRLQ 409
Db      333  FRWR-----ETLRPRSFMSAL-----AF-CKACWKLOQESRYQ 364

```

Search completed: September 9, 2004, 21:54:39  
 Job time : 36 secs

US-09-390-131-3

Query Match 9.7%; Score 255.5; DB 4; Length 451;  
Best Local Similarity 22.7%; Pred. No. 6.5e-18;  
Matches 99; Conservative 65; Mismatches 173; Indels 99; Gaps 18;

QY 35 VMELGKFERKEREKSSS-----LDDGHTXMEADPTHLNSFLKKEGLTPNKR 81  
DB 46 IYAPISKEQKDWKEGRIELSDNDELDPILQKEELKPOQRPSEFENVPKKTENFNVG 105  
QY 82 KWEIDSYPIM-----LWMSPLTGENGRIGOCGADACFTTNR 118  
DB 106 KEPFDEEVLTSDDIKLEEMNTATVPQQRKLISNAGHSQDNIG--CPDMNCEFTQVR 164  
QY 119 -----TYLHHMTKAFLEFYGTDFNIDSLPLPKAHDMVA--FHEESPKNYKLFHKPV 170  
DB 165 AAPADADAVLIAHND-----DFVPRK-----NQYVYFSGESPANGIGIPRP- 208  
QY 171 ITLFNTATFSRSHPLTQYLESTIEVLKSLRYVLPSQSKKLKRLAPLYVYGDDDP 230  
DB 209 -DYNNMTLGF-RHD-TPAGSPGYTVKLGAQSKRTGQVDAVLVNGAKAAMFVSHCQT 265  
QY 231 PSRDSYVRELMTYIEVSYECLRNKDLPQCLKNFASMDADGFYRIIAQYFIILAFENA 290  
DB 266 NSKREDPVKTLQKHLQIDYGGCPMKCARGDSKQDTMLDT-----YHFVYTFENS 317  
QY 291 VCDYITEKFWRP-LKLGVPVYVYSGPSITDMLPSKSAIIVSESHPRELASYIRLDS 349  
DB 318 ICEYVYVTEKMSGYQNTIIPVLKRLKVEFPVPPN-SFIALDPSKVMEDGYLNTYLMN 376  
QY 350 DDLRYEAYVEMKLGK-----SNQRLTLALREKRGVQDVQNDYIDAFECMVCTKYWA 404  
DB 377 NKTAYMEYFEMRDYKAVFLDGSMDVL-----ERPMGF-----CQYCRAMNT 419  
QY 405 NIRLOEKGLPPKMEWA 420  
DB 420 EPR--QKVLIP-NMWA 432

## RESULT 14

US-09-092-315-12  
; Sequence 12, Application US/09092315  
; Patent No. 6399337  
; GENERAL INFORMATION:  
; APPLICANT: Taylor, Diane E.  
; APPLICANT: Ge, Zhongming  
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE  
; FILE REFERENCE: 07254/049001  
; CURRENT APPLICATION NUMBER: US/09/092,315  
; CURRENT FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: US 60/048,857  
; EARLIER FILING DATE: 1997-06-06  
; NUMBER OF SEQ. ID NOS: 22  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 12  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Gallus gallus  
US-09-092-315-12

Query Match 9.6%; Score 251; DB 4; Length 356;  
Best Local Similarity 25.8%; Pred. No. 1.3e-17;  
Matches 84; Conservative 50; Mismatches 132; Indels 60; Gaps 12;

QY 90 IMLWMSPLTGETGRIGOC-----GADACFTTINRTYLHHMTKAFLEFYGTDFNIDSL----- 141  
DB 59 VLLWMBEP--GRPWRPADCRRRYNTIGCLLSADRG--RYGEARAVLFHHRDLALRGGLP 115  
QY 142 --PLPKAHHDMAVFHEESPKNYKLFHKPVITLNNYATGSRSHLPLTQYLSIEVL 199  
DB 116 RGPFPFRPQRWYWNWFESPSHSPGL--RGLAGLFWMTWSYRSDSDVFPYGYL----- 167  
QY 200 KSLRYLVP--LQSKVKLRLAPLYVYGDDDPDSRDSYVRELMTYIEVDYSGEGLRKXKL 259

DB 168 ----YEPSPRPFPVPRKSLVAVASWNEHAEVRYRQLKEPLIDVYG----- 215  
QY 260 PQCLKNPASMDDGFYRIIAQYFIILAFENAVCDDYTEKFWP--PLKLGVPVYVYG--SP 316  
DB 216 ---ARGMALLIEG--SVKRYSAKFTLAFENSGHDTYTEKLMKRAFAASAVPVYLGPRRA 271  
QY 317 SITDMLPSKSAIIVSESHPRELASYIRLDSDDLRYEAYVEMKLGKISNORLLALR 376  
DB 272 NYERFIIPAD-SFIHVDPPSPRLLATYKFLDKNPSYRYFAWNNKY----- 319  
QY 377 ERKMGVQDVQNDNYTDAFECMVCTYV 402  
DB 320 -----VHVTSPWDEHYCKVCEAV 337

## RESULT 15

US-07-914-281-11  
; Sequence 11, Application US/07914281  
; Patent No. 5324663  
; GENERAL INFORMATION:  
; APPLICANT: LOWE, JOHN B.  
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS  
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,  
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION  
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; ADDRESS: 1755 Jefferson Davis Highway, Fourth Floor  
; City: Arlington  
; State: Virginia  
; Country: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 19920720  
; APPLICATION NUMBER: US/07/914,281  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lavallee, Jean-Paul M. P.  
; REGISTRATION NUMBER: 31,451  
; REFERENCE/DOCKET NUMBER: 2363-060-55  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 521-4500  
; TELEFAX: (703) 486-2347  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ. ID NO.: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 374 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-07-914-281-11

Query Match 9.4%; Score 245.5; DB 1; Length 374;  
Best Local Similarity 26.6%; Pred. No. 5.4e-17;  
Matches 94; Conservative 42; Mismatches 117; Indels 101; Gaps 19;

QY 90 IMLWMSPLTGETGRIGOC-----GADACFTTINRT-----YLHHMTKAFLEFYGTDFN 137  
DB 78 ILLMTWPTNTPYA-LPRSEWVPGADCNITADSQVYPQADAVIVHWD---IMYNSAN 133  
QY 138 IDSPLPKAHHDMAVFHEESPKNYKLFHKPVITLNNYATGSRH-----SHLPLTQY 192  
DB 134 LP--PPTSPQGRWTFWFMESPSN-----CRHLEALDGYFNLTMYSY 172  
QY 193 LESIEVLKSLRYLV-----LQSKVKLRLAPLYVYGDDDPDSRDSYVREL 241

```
Qy 302 R-PLKGVPPVYSGPSITDM---LPSNKSAILVSEFSHPRELASYIRLDDRLYKAYV 357
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 RNALFAAVVVLG-PERSYVERFLPD-AFIHVDPQSPKRLARLQELDDHARLYSY 318
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 358 VEWTKGEISNORLLTLRER--KMGVQDVQNDYIDAEFCVCTKWANILQSKGL 413
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 FRMR-----ETLRPRSPSMAL-----AF-CACWKLQESRYQTRGI 354
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-09-390-131-8
; Sequence 8, Application US/09390131
; Patent No. 6461835
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Cummings, Richard D.
; APPLICANT: Nyame, A. Kwame
; APPLICANT: DeRose-Boyd, Russell A.
; TITLE OF INVENTION: FUCOSYLTRANSFERASES, POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING FUCOSYLTRANSFERASES, AND TRANSGENIC MAMMAL
; FILE REFERENCE: 6679.US.01
; CURRENT APPLICATION NUMBER: US/09/390,131
; CURRENT FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-390-131-8

Query Match 9.7%; Score 255.5; DB 4; Length 393;
Best Local Similarity 26.1%; Pred. No. 5.2e-18;
Matches 95; Conservative 52; Mismatches 138; Indels 79; Gaps 16;

Qy 90 IMLWSP-----LTGETGRLGCGGACGFTINRTYLHHTKAF 129
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 VLLMWPGRWRPADCRNRNITGCLLSADRGYGA-----KAVLFHHDLA- 106
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 130 LFGYTDENIDSLPLPKRAHDMAVFHEESPKNYKLFHKVITLENVYATFSRSHLPLT 189
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 107 -LHGRO-GLPRGPPRPGRQWMMNFESPSSHSGL--RGLAGLFMTWSYRSDSDVFPV 162
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 190 TQYLESIVLSLVIYVLOSNNKLARKLAPLYVVOGSDCPSPDRGSYVREIMTYEVS 249
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 163 YGYL-----YEPSPRPFPVLPKRSRLVAVVISNNNEHARVRYTQLKEHLPIV 212
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 250 YGECRLNKLQOLKNPASMADGFRYIIAQYKFIILAFENAVDDYITEKFWR-PLKLG 308
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 YG-----ARGMALLLEG-SVVKTVSAKKFYLAFFENSQHTDYITELWMNAFAAGA 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 309 VPVYVG--SPSITDMLPSNKSAILVSEFSHPRELASYIRLDDRLYKAYVEMKLGSI 366
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 VPVVLGPPRRAYVERIPAD-SFIHVDPFSPRLIATYLKELDKKSSYRRYFAMRNKYE 319
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 367 SNORLLTLRER-----RKMGVQDVQNDYIDAEFCVCTKWANIRLQEKGLPKR 416
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 H-----VTFMWEHVKCKVEAVRTGNQKTYQNLAGNFE-SVCRMWMTERR--QKVLIP- 371
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 417 RWEA 420
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 NWDA 375

RESULT 12
US-09-390-131-9
; Sequence 9, Application US/09390131
; Patent No. 6461835
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Cummings, Richard D.
; APPLICANT: Nyame, A. Kwame
; APPLICANT: Nyame, A. Kwame
US-09-390-131-9
```

```
; APPLICANT: DeRose-Boyd, Russell A.
; TITLE OF INVENTION: FUCOSYLTRANSFERASES, POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING FUCOSYLTRANSFERASES, AND TRANSGENIC MAMMAL
; TITLE OF INVENTION: INCORPORATING SAME
; FILE REFERENCE: 6679.US.01
; CURRENT APPLICATION NUMBER: US/09/390,131
; CURRENT FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-390-131-9

Query Match 9.7%; Score 255.5; DB 4; Length 450;
Best Local Similarity 22.7%; Pred. No. 6.5e-18;
Matches 99; Conservative 65; Mismatches 173; Indels 99; Gaps 18;

Qy 35 VMEVLKFEKFEKSS-----LDDGHTKMEAFTHLSLKKEGLTENKR 81
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 46 IYAPISKSECKWKEGIELSNDHELDPFLOKEBLKQRPSEENVPKKTFFNFVVG 105
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 82 KWEIDSYPIW-----LWSPPLTGTGRLGCGGACGFTINR 118
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 106 KEPPDVEVLTSSDIKBERMTATVPQCKRLLSMNGHSDNLQG-CPMNCEFTQVR 164
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 119 -----TYLHHTMTAFLFYGTDENIDSLPLPKRAHDMAV-FHEESPKNYKLFHKV 170
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 ARAPDDAVLIAMDN-----DVPKPR--NQYVYFSGESPANSGIQIPSP- 208
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 171 ITLFNTATPSRSHLPLTQYLESIEVLSKSLRYVLPLOSNNKLARKLAPLYVVOGSDCP 230
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 209 -DYINMTLGF-RND-TPAGSPGYTYKLGAKSRKTQGVVDALVNGKAGAMFVSHQT 265
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 231 PSRDSYVREIMTYEIVSDYECRLNKDLPQOLKNPASMADGFRYIIAQYKFIILAFENA 290
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 NSKREDPVKKLQHLQIDIVGGCGPWKCARCDKCDTMDTD-----YHFTVFENS 317
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 291 VCDYITEKFWR-LKGVPPVYSGPSITDMLPSNKSAILVSEFSHPRELASYIRLDS 349
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 318 ICEDYITEKMKSGYQVITLPLVKRKLYEPVFPN-SLIAIDCKSVKEMQDIANYLMN 376
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 350 DDRLYKAYVEMKLGSI-----SNORLLTLRERKMGVQDVQNDYIDAEFCVCTKWYA 404
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 377 NKTAVMYEFWRMDYKVFVLDGSHDVL---ERPWF-----CQVRMAWT 419
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 405 NIRLQEKGLPKKWEA 420
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 420 EPR--QKVLIP-NWDA 432
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-09-390-131-3
; Sequence 3, Application US/09390131
; Patent No. 6461835
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Cummings, Richard D.
; APPLICANT: Nyame, A. Kwame
; APPLICANT: DeRose-Boyd, Russell A.
; TITLE OF INVENTION: FUCOSYLTRANSFERASES, POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING FUCOSYLTRANSFERASES, AND TRANSGENIC MAMMAL
; FILE REFERENCE: 6679.US.01
; CURRENT APPLICATION NUMBER: US/09/390,131
; CURRENT FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-390-131-3
```

```

; APPLICATION NUMBER: US/08/393,246
; FILING DATE:
; APPLICATION NUMBER: US 08/220,433
; FILING DATE: 30-MAR-1994
; APPLICATION NUMBER: US 07/914,281
; FILING DATE: 20-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavallee, Jean-Paul M. P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-09-042-531-14

Query Match 10.0%; Score 263; DB 3; Length 359;
Best Local Similarity 27.4%; Pred. No. 7,4e-19;
Matches 98; Conservative 48; Mismatches 118; Indels 94; Gaps 19;

QY 87 SYPIIMWMSPLTGETGRIGQC-----GADACFFITNR-----TYLHHMTKAFLEYG 133
DB 60 SIPLILMTWPFNKRIALPRGSEWPGTADCNITADRKVYPOADAVIVHRE-----VMYN 115
QY 134 TDENIDSLP-LPRKAHDMAVFHEESPKNKYKLFHKPVITLFPNTATFSRSHSLPLTQY 192
DB 116 PS---AQLPRSPRQOGQRIWIFSMESPSHCWL--KAMDGYFN-----LTMYSY 158
QY 193 LESIEVLKSLRYLV-----LQSKKKLRKRLAPLVYVQSDCPPSDRDSYREL 241
DB 159 RSDSDIFTPYGMLEPWSGQPAHPPLNLSAKTEL-----VAMAVSNMGPN SARVRYOQL 212
QY 242 MTYIEVDSYGECLEKRNKDLPOQLKNPASMADGFRITIAQYKFLAFENAVCDYITEKFW 301
DB 213 QAHLEKVDVYGR--SHKPLPQ-----GTMETLSRYKFLAFENSLHPDYITEKLM 260
QY 302 R-PLKLGVPVYVYGSPTIDW---LPSNKSALIVSEFSPRELASYIRRLDSDRLYEAY 357
DB 261 RNALAMAVPVVLG-PSRSNYERFLPPD-AFIHVDDFQSPDLARYLQELDKDHARLYSY 318
QY 358 VEMKLGKGEISNQLRLTLALRER--KMGVDVNDVNDYIDAFEGMCTKYMANIRLQEKGL 413
DB 319 FRWR-----ETLRPRSFQWAL-----AF-CKACWKLGQESRYQTGRI 354

RESULT 9
; US-09-092-315-10
; Sequence 10, Application US/09092315
; Patent No. 6399337
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/09/092,315
; CURRENT FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: US 60/048,857
; EARLIER FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-092-315-10
Query Match 10.0%; Score 263; DB 4; Length 359;
```

```

Best Local Similarity 27.4%; Pred. No. 7,4e-19;
Matches 98; Conservative 48; Mismatches 118; Indels 94; Gaps 19;

QY 87 SYPIIMWMSPLTGETGRIGQC-----GADACFFITNR-----TYLHHMTKAFLEYG 133
DB 60 SIPLILMTWPFNKRIALPRGSEWPGTADCNITADRKVYPOADAVIVHRE-----VMYN 115
QY 134 TDENIDSLP-LPRKAHDMAVFHEESPKNKYKLFHKPVITLFPNTATFSRSHSLPLTQY 192
DB 116 PS---AQLPRSPRQOGQRIWIFSMESPSHCWL--KAMDGYFN-----LTMYSY 158
QY 193 LESIEVLKSLRYLV-----LQSKKKLRKRLAPLVYVQSDCPPSDRDSYREL 241
DB 159 RSDSDIFTPYGMLEPWSGQPAHPPLNLSAKTEL-----VAMAVSNMGPN SARVRYOQL 212
QY 242 MTYIEVDSYGECLEKRNKDLPOQLKNPASMADGFRITIAQYKFLAFENAVCDYITEKFW 301
DB 213 QAHLEKVDVYGR--SHKPLPQ-----GTMETLSRYKFLAFENSLHPDYITEKLM 260
QY 302 R-PLKLGVPVYVYGSPTIDW---LPSNKSALIVSEFSPRELASYIRRLDSDRLYEAY 357
DB 261 RNALAMAVPVVLG-PSRSNYERFLPPD-AFIHVDDFQSPDLARYLQELDKDHARLYSY 318
QY 358 VEMKLGKGEISNQLRLTLALRER--KMGVDVNDVNDYIDAFEGMCTKYMANIRLQEKGL 413
DB 319 FRWR-----ETLRPRSFQWAL-----AF-CKACWKLGQESRYQTGRI 354

RESULT 10
; US-09-733-524A-10
; Sequence 10, Application US/09733524A
; Patent No. 6534298
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING ALPHA-1,3
; TITLE OF INVENTION: FUCOSYLTRANSFERASES AND EXPRESSION SYSTEMS FOR MAKING AND
; FILE REFERENCE: 07254-049002
; CURRENT APPLICATION NUMBER: US/09/733,524A
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 09/092,315
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-733-524A-10

Query Match 10.0%; Score 263; DB 4; Length 359;
Best Local Similarity 27.4%; Pred. No. 7,4e-19;
Matches 98; Conservative 48; Mismatches 118; Indels 94; Gaps 19;

QY 87 SYPIIMWMSPLTGETGRIGQC-----GADACFFITNR-----TYLHHMTKAFLEYG 133
DB 60 SIPLILMTWPFNKRIALPRGSEWPGTADCNITADRKVYPOADAVIVHRE-----VMYN 115
QY 134 TDENIDSLP-LPRKAHDMAVFHEESPKNKYKLFHKPVITLFPNTATFSRSHSLPLTQY 192
DB 116 PS---AQLPRSPRQOGQRIWIFSMESPSHCWL--KAMDGYFN-----LTMYSY 158
QY 193 LESIEVLKSLRYLV-----LQSKKKLRKRLAPLVYVQSDCPPSDRDSYREL 241
DB 159 RSDSDIFTPYGMLEPWSGQPAHPPLNLSAKTEL-----VAMAVSNMGPN SARVRYOQL 212
QY 242 MTYIEVDSYGECLEKRNKDLPOQLKNPASMADGFRITIAQYKFLAFENAVCDYITEKFW 301
DB 213 QAHLEKVDVYGR--SHKPLPQ-----GTMETLSRYKFLAFENSLHPDYITEKLM 260
```